DOCUMENT RESUME

ED 148 470 PS 009 691

AUTHOR	Lazar, Irving; And Others
TITLE	The Persistence of Preschool Effects: A Long-Term Follow-Up of Fourteen Infant and Preschool Experiments. Final Report.
INSTITUTION	Education Commission of the States, Denver, Colo.; State Univ. of New York, Ithaca. Coll. of Human Ecology at Cornell Univ.
SPONS AGENCY	Administration for Children, Youth, and Families (DHEW), Washington, D.C.
PUB DATE	Sep 77
GRANT	OCD-18-76-07843
NOTE	221p.; For summary to this report, see PS 009 692
EDRS PRICE	MF-$0.83 HC-$11.37 Plus Postage.
DESCRIPTORS	Academic Achievement; Child Care Centers; *Early Childhood Education; *Educational Research; *Followup Studies; Home Programs; *Longitudinal Studies; Low Income Groups; Parent Education; *Program Effectiveness; *Program Evaluation; Research Methodology
IDENTIFIERS	*Developmental Continuity

ABSTRACT

This report summarizes the findings of the Developmental Continuity Consortium, a collaborative effort of twelve research groups conducting longitudinal studies on the outcomes of early education programs for low-income infants and preschool children initiated in the 1960's. The educational experiments conducted were of three general types: (1) home-based parent education programs; (2) center-based preschool programs; and (3) combined preschool center and home visit programs. Data discussed in this report include (1) data collected independently by each research group at the beginning of its program and over a number of years after the program ended; and (2) follow-up data collected in 1976-77 when the subjects were 9-18 years old. Measures used in the current follow-up study include Parent and Youth Interviews, IQ tests, achievement tests, and data from school records. Data were collected for program children and their control or comparison group. Combined results indicate that (1) early education significantly reduced the number of children assigned to special classes; (2) early education reduced the number of children held back one or more grades; and (3) children from all three types of programs surpassed their controls on IQ tests for up to three years after the end of the program. Results, conclusions and implications, methodological problems and further analyses are discussed. (SB)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC make
effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this aff
quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (
EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be ma
the original.

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE PERSISTENCE OF PRESCHOOL EFFECTS

ED148470

PS 009691

A NATIONAL
COLLABORATIVE STUDY BY
TWELVE RESEARCH GROUPS
FOR THE
EDUCATION COMMISSION OF THE
STATES. SEPTEMBER, 1977

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Irving Lazar

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

COMMUNITY SERVICE LABORATORY
NEW YORK STATE COLLEGE OF HUMAN ECOLOGY
CORNELL UNIVERSITY ITHACA, N.Y. 14853

THE PERSISTENCE OF PRESCHOOL EFFECTS

A Long-Term Follow-Up of Fourteen Infant and Preschool Experiments

The Consortium on Developmental
Continuity
Education Commission of the States

Analysis and Final Report by
Irving Lazar, Virginia Ruth Hubbell,
Harry Murray, Marilyn Rosche, and
Jacqueline Royce

PS 009691

Final Report; Grant No. 18-76-07843
To The Administration on Children, Youth and Families
Office of Human Development Services
U.S. Department of Health, Education and Welfare
September, 1977

Members/Collaborators of the Consortium on Developmental Continuity

Kuno Beller
Temple University

Ira Gordon*
University of Florida

Martin & Cynthia Deutsch
New York University

Susan Gray
George Peabody College

Merle Karnes
University of Illinois

Irving Lazar (Chairman)
Cornell University

Phyllis Levenstein
Verbal Interaction Project

Louise Miller
University of Louisville

Francis Palmer
SUNY - Stony Brook

David Weikart
High Scope Foundation

Myron Woolman
Institute for Educational Research

Edward Zigler
Yale University

Central Staff of the Consortium (at Cornell University)

Virginia Ruth Hubbell

Harry Murray

Jane Pedersen

Marilyn Rosche

Jacqueline Royce

Project Director

James Peterson
Education Commission
of the States
Denver, Colorado

Principal Consultants

Bernard Brown (ACYF)

Richard Darlington (Cornell)

Robert McCall (Boys Town Inst.)

---

*Dr. Gordon is now at the University of North Carolina at Chapel Hill.

This report was prepared by the Chairman and the Staff of the Consortium. Inquiries should be addressed to The Chairman at N-135 MVR Hall, Cornell University, Ithaca, New York 14853.

It would be impractical to acknowledge all of the people whose help made this complex collaboration work so well as it has. We wish to particularly thank some of those who made important contributions to the quality of our work. While the remaining faults are ours, they deserve some of the credit for its strengths.

Edith Grotberg and Bernard Brown of the Administration for Children, Youth and Families (DHEW) helped us think through critical issues in selecting our variables.

James Peterson and Homer Elseroad at the Education Commission of the States opened the doors for cooperation by public school systems across the country.

Richard Darlington, Urie Bronfenbrenner, John Doris, Henry Ricciuti, Steven Caldwell and Jay Millman of Cornell, and Robert McCall of Boys Town, served as consultants, critics, and advisors throughout the process of selecting data treatment and interpretation alternatives.

The research associates of the members of the Consortium worked well beyond their formal obligations to find and gather this enormous body of data.

TABLE OF CONTENTS

PART I. INTRODUCTION

A. Background 1

B. Formation of Consortium 5

C. Description of Studies 7

D. Consortium Problems, Decisions 14

E. Problems of Secondary Analysis 21

F. Problems of Longitudinal Analysis 22

G. Overview of the Analysis 25

PART II. DATA COLLECTION METHODOLOGY

A. Original Data 27

B. Follow-up Data 28

PART III. COMPARISON OF SAMPLES ACROSS PROJECTS 35

PART IV. ANALYSIS OF PROGRAM AND CONTROL DIFFERENCES IN COGNITIVE FUNCTIONING 45

A. Analysis of Stanford-Binet IQ Scores from Immediate Posttest to Three Years after Program Termination 47

B. Analysis of WISC-R IQ Scores 58

C. Summary of Results of the IQ Testing 61

D. Analysis of Assignment to Special Education and Grade Failure 62

E. Analysis of Schooling Variables from Youth Interview 82

PART V. ANALYSIS OF DELIVERY SYSTEMS

A. Parental Evaluations of Programs 85

B. Program Variables Analysis 91

PART VI. CONCLUSIONS AND IMPLICATIONS

A. Conclusions 107

B. Implications and Recommendations for National Policies and Programs 109

PART VII. FUTURE TASKS

A. WISC-R 112

B. Achievement Tests 112

C. Special Education and Retention 113

D. Interview Responses 113

E. Who Benefits from What, When, and for How Long? 113

F. Developmental Questions 114

PART VIII. APPENDICES

A. Status Report

1. Introduction A-1

2. Consortium Meetings A-2

3. Presentations Given 1976-77 A-4

4. Status by Instrument A-6

B. Technical Supplements

1. Sample Selection and Assignment B-1

2. Follow-up Data Collection and Data Analysis Methods (Reliability, Response Rates) B-20

3. Attrition B-33

4. Relationship of Current Results on Special Education and Grade Failure to "Preliminary Findings of the Developmental Continuity Longitudinal Study" B-74

5. Glossary B-76

6. Instruments and Instructions

C. Papers Presented

LIST OF TABLES

Table 1: Summary of Studies 13

Table 2: Comparison of Projects on Pretest Stanford-Binet IQ 39

Table 3: Comparison of Projects on Mother's Education 41

Table 4: Comparison of Projects on Head of Household SES 43

Table 5: Mean IQ's for Experimental and Control by Project 54

Table 6: Comparing Program and Control on Stanford-Binet IQ Score - Pooled z's and Significance for t-tests by # Years after Termination of Early Education 55

Table 7: Comparison of t-test results with Regression on Effect of Program 56

Table 8: Pooled z's Substituting in Regression Coefficient Where Available 57

Table 9: Comparison of WISC IQ's by Project 60

Table 10: Effects of Early Education on Special Education for Five Projects 74

Table 11: The Effect of Different Curricula on Assignment to Special Education 76

Table 12: Effects of Early Education on Grade Failure for Five Projects 77

Table 13: The Effect of Different Curricula on Grade Failure 79

Table 14: The Effects of the Micro-Social Learning System on Grade Failure 80

Table 15: The Effects of the Micro-Social Learning System on Assignment to Special Education 81

Table 16: Self-evaluation of School Work by Program or Control Status (older project sites only) 83

Table 17: Responses to Parental Interview Question "Was the program a good thing for your child?" By Delivery System 86

Table 18: Answers to Question "What did you like best about the program your child was in?" By Delivery System 87

Table 19: Responses to Question "Did you like having the program in a home/center?" By Delivery System 89

Table 20: Correlations between Immediate Posttest IQ and Treatment-Control Status 96

Table 21: Matrix of Program Descriptions 97-104

Table 22: Z-Score Computation for Between Program Variable Categories 105-106

Table A-1: Number of Cases Received by September 30, 1977 for Each Instrument by Project Site A-9

Table A-2: Number of Cases in July 1, 1977 Computer File for Each Instrument by Project Site A-10

Table B-1: Primary Source of Subjects B-17

Table B-2: Sample Selection Criteria Used by Projects B-18

Table B-3: SES Selection Criteria B-19

Table B-4: Response Rate Computations and Final Dispositions for Six Sites B-30

Table B-5: Final Sample as a Percent of Original Sample by Project B-47

Table B-6: Final Sample as a Percent of Original Sample of Parent Interview, by Project B-48

Table B-7: Final Sample as a Percent of Original for Youth by Project B-49

Table B-8: Final Sample as a Percent of Original for School Record Form by Project B-50

Table B-9: Final Sample as a Percent of Original for WISC-R by Project B-51

Table B-10: General Attrition - Comparison of Pretest Stanford-Binet IQ Scores B-52

Table B-11: General Attrition - Comparison of Hollingshead ISP Scores B-53

Table B-12: General Attrition - Comparison of Mother's Education B-54

Table B-13: Attrition on the Parent Interview - Comparison of Pretest Stanford-Binet IQ Scores B-55

Table B-14: Attrition on the Parent Interview - Comparison of Hollingshead ISP Scores B-56

Table B-15: Attrition on the Parent Interview - Comparison of Mother's Education B-57

Table B-16: Attrition on the Youth Interview - Comparison of Pretest Stanford-Binet IQ Scores B-58

Table B-17: Attrition on the Youth Interview - Comparison of Hollingshead ISP Scores B-59

Table B-18: Attrition on the Youth Interview - Comparison of Mother's Education B-60

Table B-19: Attrition on the School Record - Comparison of Stanford-Binet IQ Scores B-61

Table B-20: Attrition on the School Record - Comparison of Hollingshead ISP Scores B-62

Table B-21: Attrition on the School Record - Comparison of Mother's Education B-63

Table B-22: Attrition on the WISC-R - Comparison of Pretest Stanford-Binet IQ Scores B-64

Table B-23: Attrition on the WISC-R - Comparison of Hollingshead ISP Scores B-65

Table B-24: Attrition on the WISC-R - Comparison of Mother's Education B-66

Table B-25: Expected Number of Chance "Significant Findings" and Actual Number of Significant Findings for Tables B-10 to B-24 B-67

Table B-26: Results of 3 Way ANOVA: Significance Levels B-68

Table B-27: Attrition on the Original Data: Comparison of Stanford-Binet Pretest IQ Scores B-69

Table B-28: Attrition on the Original Data: Comparison of ISP Scores B-70

Table B-29: Attrition on the Original Data: Comparison of Mother's Education B-71

LIST OF FIGURES

Figure 1: Homogeneous Groups of Projects on Pretest Stanford-Binet IQ for the Original Sample 40

Figure 2: Homogeneous Groups of Projects on Pretest Stanford-Binet IQ for the Follow-up Sample 40

Figure 3: Homogeneous Groups of Projects on Mother's Education for the Original Sample 42

Figure 4: Homogeneous Groups of Projects on Mother's Education for the Follow-up Sample 42

Figure 5: Homogeneous Groups of Projects on Socio-economic Status for the Original Sample 44

Figure 6: Homogeneous Groups of Projects on Socio-economic Status for the Follow-up Sample 44

Figure 7: Percent of Program and Control Children in Special Education 75

Figure 8: Percentage Reduction in Children in Special Education 75

Figure 9: Percent of Program and Control Children Held Back a Grade 78

Figure 10: Percentage Reduction in Children Held Back in Grade 78

The Persistence of Preschool Effects

First Year Report

Abstract

I. Overview

The Developmental Continuity Consortium is a collaborative effort of twelve research groups conducting longitudinal studies on the outcomes of early education programs. This report summarizes the findings of current analyses of longitudinal studies of low income children who participated in experimental infant and preschool programs initiated in the 1960's.

II. Methods

These educational experiments were of three general types: (1) homebased, parent education-infant programs, (2) preschool center programs, and (3) combined preschool center and home visit programs. The two sets of data discussed in this report include 1) data collected independently by each research group at the beginning of their program and over a number of years after the program ended; and 2) common follow-up data collected in 1976-77. These low-income youngsters, who are predominately black, now range from nine to eighteen years of age. Parent and Youth Interviews, Wechsler IQ scores, achievement test scores, and data from school records were collected in the current follow-up on program children and their control or comparison group. Methodological problems inherent in longitudinal research and secondary analysis such as attrition and sample selection variations are discussed. Strategies for investigating program characteristics (age and length of intervention, curriculum types) are proposed for future analyses.

III. Findings

The results analyzed thus far show that investments in early education have long term benefits in three areas:

1) Special Education Assignment. The combined evidence from project sites which collected this data shows that early education significantly reduced the number of program children assigned to special educational classes.

2) In-grade retention. The combined evidence from seven project sites able to collect this information indicates that early education significantly reduced the number of children held back one or more grades.

3) Cognitive measures. The children from all three types of programs surpassed their controls for up to three years after the end of the program on the Stanford-Binet. This significant difference appears to last through the primary grades. Current Wechsler results show only the youngest program subjects with IQ scores significantly higher than their controls'.

Attrition analyses indicate that over all, there were no significant differences between the early demographic and cognitive characteristics of those subjects found and those not found for follow-up study. Less than 3% of those contacted refused to participate in the follow-up study.

IV. Conclusions

The most important finding is that low income children who received early education are better able to meet the minimal requirements of their schools as shown in reduced rate of assignment to special education and in-grade retention. In addition, the preschool programs improve cognitive skills into the primary grades as measured by the Stanford-Binet. Parental satisfaction with programs was high as measured on the follow-up interviews. Preliminary analyses indicate significant differences in attitudinal responses such as program children rating themselves better than others in their school work compared to their controls.

V. Implications

The implications are that well planned curricula for young children in day care and Head Start are likely to reduce later costly special education or remedial programs in schools.

PART I

INTRODUCTION

A. Background

Alexis de Tocqueville, writing at a time when broad generalizations were still respectable, observed that Europeans tended to accept their stations in life and to try to improve conditions within their stations, whereas Americans assumed that they could change their station through hard work, virtue and talent. While in recent years we have added the idea of opportunity, de Tocqueville's generalization about Americans can be defended today.

Head Start was built on the assumption that early education, parental involvement, and the provision of medical and social services could enable children of poor parents to achieve cognitive parity with their middle-class peers, and leave the ranks of the poor.

From this traditional American premise, a vast array of programs and curricula emerged. The age of entry into programs, the length of intervention, the settings, the teachers, and the materials employed all varied in a profuse diversity of programs for low-income children and their families.

Twelve years later, with sufficient numbers of program children in their later years of childhood and adolescence, it became appropriate to take stock of outcomes.

This report describes the findings of fourteen longitudinal studies of low-income children who participated in experimental infant and preschool programs prior to 1969. It was conducted as a collaborative effort by a dozen investigators who agreed to collect common follow-up data in 1976-77.

We believe that the findings in this report now leave no reasonable doubt that in the main, programs which had deliberate cognitive curricula had a significant long-term effect on school performance. The findings are both too conservative in the methods of data analysis from which they emerge and too dramatic in their consistency and size, for the main effects they found to be spurious. The critical questions they address could, we believe, be addressed only by longitudinal methods.

Longitudinal data provide the most valid and direct way of assessing the cognitive, social, emotional and familial outcomes of programs for young children. Scientifically useful longitudinal data are also the most difficult to obtain, and longitudinal studies are the most difficult to design and maintain.

Intervention programs typically have small N's; our mobile society produces high attrition; changes in psychometric technology make baseline measures obsolete; changing political climates can suddenly cut off access to data or research funds. Further, policy-makers are rarely content to wait for findings -- but children's growth cannot be hurried by electoral calendars.

The demonstration that investments in early education have long-term benefits that are both humane and fiscal cannot help but quell the doubts that have effectively frozen such investments for almost a decade.

Theoretical Significance

In recent years, there has been a resurgence of the perennial populist-elitist debate about the malleability of the child's mind. It is perhaps symbolic that today's most eloquent champions of the elitist position speak from positions at our oldest -- and traditionally, most elite -- university. It is from Harvard that we hear that intelligence is primarily genetic (Herrnstein); that anything that can be done must be done before three years of age -- and then by a supermom who stays home (White); that early trauma will be "outgrown" anyway, without the taxpayer's investments (Kagan); and that, in any event, if a poor kid makes it, it's purely a matter of luck (Jencks). This modern echo of Cotton Mather is also part of the American experience, and only the clearest practical demonstration that talents, learning ability, and capacity for social development are widely distributed, and can be actualized by appropriate stimulation, will permit continued national investment in Head Start's role in the search for ways of increasing human capabilities. The ideological and theoretical arguments are not new. What is new is the opportunity to move part of the debate from the realm of speculation to scientific test. That opportunity exists because of the investments in early intervention of the sixties, and OCD's decision to risk the outcomes of longitudinal studies.

OCD has been no stranger to longitudinal studies. Where other federally supported longitudinal research usually did not involve specific interventions (Perinatal Study, Children of Kauai, Berkeley Growth Study), the ETS Longitudinal Study, the follow-ups of children in the PCC's and the PCDC's, the current studies of Home Start, and this research all have the potential of testing the developmental hypothesis; of reaching new insights into the relationship between early experience, the conditions and settings of a young child's life, and the behavioral, social and educational outcomes of planned early intervention.

The reader is asked to keep a few caveats in mind as (s)he reads this report.

(1) While the curricula and delivery systems used in these experiments can be found in Head Start programs in many places and could be easily adopted by others, and while the children were typical of Head Start's populations, these were not typical Head Start programs. They were experimental programs. They varied in ages, frequency and duration of sessions. However, some were actually Head Start sponsored, and current Head Start quality standards are such that similar curricula are likely to be part of typical Head Start programs.

(2) These studies, with a few exceptions, were not initially designed as longitudinal studies, nor were they designed for later comparisons or pooling of the data. This is a secondary analysis, and there are very real limits on the amount of information that was common across studies when they collected their initial baseline information. However, they are, with a few exceptions, all the existing studies that can be used for the investigation of long-term effects of early intervention. It would take another fifteen years (and at least five million dollars) to create a similar sample.

In this report we describe in detail each of the analyses and their limitations so that technical readers can assess our methods and conclusions for themselves. Some of the questions for which we now have the data will need to be dealt with in future reports. This report includes in its analyses raw data received up until the first week of July, 1977. More data have been received since then, and the quantities of such additional data are indicated. Full information on all located subjects will probably not arrive until late in October.

Not all comparisons include all project sites. This is due, not to deliberate selection on our part, but to the availability of the data at time of analysis.

B. Formation of the Consortium

The idea of this Consortium arose out of at least four interests that converged in 1975.

- Several of the investigators were very much concerned with finding ways to follow their original subjects, primarily for scientific reasons.
- There was a building band-wagon of professionals who were denying the importance of early experience or continuity of experience in later life - an idea which, to many other professionals, not only defied common observations, but was based upon what they felt was very flimsy and highly selective evidence.
- There were clear indications, particularly in published budget projections, that the Office of Management and Budget sought to severely reduce federal expenditures for young children. The virtual abolition of all but custodial requirements for federally supported day care seemed a harbinger of the future.
- A number of state governments were seeking help in selecting a sensible policy in regard to early childhood programs.

During 1975 a number of the investigators who were attending a national meeting met informally and agreed that the only proper "test" of the effects of early experience was to seek out children who had been subjects in early intervention studies long enough ago for any effects to have appeared. The early intervention studies of the late 1950's and early 1960's seemed a fruitful source of such persons.

Irving Lazar, of Cornell University, who was then examining the research issues relating to the continuity of development for the

Social Research Group at George Washington University and serving as an advisor on program development for the Education Commission of the States, was able to obtain some travel funds to call together persons who had directed early intervention programs which had certain basic characteristics:

- They had completed their studies prior to 1969.
- They had large enough original samples so that a recovery rate of 30% would still be a statistically usable number.
- They met some basic conditions of experimental design.
- They had made some effort to follow their subjects.
- They had explicit and standard intervention programs, so that the content of the child's experience could be specified.

At the initial meeting, sponsored by the Social Research Group, two other investigators with sizable numbers of subjects were nominated for inclusion in the proposed collaborative arrangement. One of these accepted the invitation; the other made no reply, and since his data had not been made available to the scientific community, no further pursuit was undertaken.

Because of the importance of the proposed follow-up to state governments, the Education Commission of the States, which serves as a policy research organization for forty-nine state governments, offered to sponsor the Consortium effort.

At a second meeting of the Consortium group, certain basic agreements were undertaken:

(1) Each member would turn over a set of his/her original data and follow-up data for central processing to a group who had not been involved in any of the studies.

(2) Each member would retain "ownership" of those data, in the sense that he/she could review in advance the uses which the central group would make of them and retained full rights to publish his/her own findings. The Consortium could use the data for its common purposes, but only individual investigators could release the raw data to others.

(3) Each member agreed in advance to collect a common set of protocols, to be jointly arrived at, in the new follow-up.

It was declared that Dr. Lazar would chair the Consortium, would direct the data treatment at Cornell, and would take the lead in seeking financial support for the studies.

C. Description of Studies

The intervention studies included in this report were designed to determine whether an individual program with children had durable results. They involved treatments which varied as a function of the theoretical disposition of the principal investigator. Age, type, duration, and intensity of intervention varied from study to study. Most of the programs were begun in the early and middle 1960's. Federal funding was obtained from various sources, including the Office of Economic Opportunity, the U.S. Office of Education, the Children's Bureau, the National Institute of Mental Health, and the National Institute of Child Health and Human Development. Additional support came from private funds, notably from the Carnegie Corporation and Ford Foundation. While each of the intervention studies was originally designed to determine whether a particular treatment (or treatments) was effective, as a group they provide an answer to whether or not intervention as a concept is effective.

The studies from which the data presented here were drawn may be grouped by delivery system into three categories. Home-based delivery systems direct their educational efforts primarily toward the parent, usually the mother, as the major instrument of change and influence in the child's life. Activities, toys, and games are brought to the family home by a "home visitor" who trains the mother to use the activities and to promote her child's development through parent-child interaction. Center-based programs provide more or less structured nursery school programs for children. Interaction usually occurs in small groups but in some cases is on a one-to-one basis. Parents are kept informed about the program but are not actively involved in day-to-day intervention. Each combination program consists of a center-based nursery school program coupled with a periodic home visit in which both parent and child are involved.

The characteristics of the studies included in this research are summarized in Table 1. Below is a brief description of each of the programs. Additional information, such as the details of sample selection, are included elsewhere in this report.

The Philadelphia Project: Dr. Kuno Beller

Dr. Beller's program studied the effects of variations in the timing of entrance into preschool or school. Three groups of children were involved. An experimental nursery school was provided through the school system for a group of four year old children. Classes of fifteen children, attended by a head teacher and an assistant, operated four days a week. A second group of children began regular public school kindergarten at age five, and a third group entered school in the first grade at age six with no kindergarten preschool experience.

Institute for Developmental Studies: Drs. Martin and Cynthia Deutsch

The Deutsch's program examined the effects of a specific intervention program on several sample waves of children from low-income areas in New York City. They compared each of these to three control groups of children from the same areas. Their specifically developed curriculum, which began with a preschool program and extended into elementary school through third grade, emphasized language development, concept formation, perceptual and overall cognitive development, and the child's self-concept.

The Parent Education Program: Dr. Ira Gordon

Dr. Gordon's project provided home-visitor, parent-focused intervention during the middle 1960's to children from three months to three years old. This study was specifically focused on the enhancement of the intellectual and personality development of the child and the production of changes in the mother's self-esteem and in her conviction that she could affect what happened to herself and her child. Gordon utilized trained paraprofessional home visitors who worked with each mother once a week. The sequenced curriculum emphasized Piagetian concepts appropriate to the child's stage of development. One treatment group received weekly visits for two years, starting when the child was three months old; a second, visits from three months to one year of age; and a third, visits from one year to two years of age.

As the children reached two years of age, they entered Gordon's Home Learning Center Program for an additional year. This treatment continued weekly home visits to the parent, but added a twice-weekly four hour group experience for five children at a time. The "backyard centers" were located in the homes of the families in the project.

Early Training Project: Dr. Susan Gray

Dr. Gray's program utilized a single intervention model but implemented it through two delivery systems, a center-based summer program and a home visitor winter program. The project was directed towards developing the child's attitudes and aptitudes conducive to school success, as well as his or her general competence, and towards encouraging the parent to become a more effective teacher of the child.

Family-Oriented Home Visitor Program: Dr. Susan Gray

This home-based program focused on enabling the parents to become more effective educational change agents with their small children. Home visitors worked with the mother and toddler plus one other child to improve the mother's effectiveness as an educational change agent. This took place weekly for eight months. The entire family was involved wherever possible.

Curriculum Comparison Study: Dr. Merle Karnes

In this study, each of five groups of preschool children attended programs offering different curriculum models: Bereiter-Engelmann, traditional, Community-Integrated, Montessori, and Dr. Karnes' concept development curriculum. Each group attended one of the preschool models for about two hours a day for seven to eight months.

Mother-Child Home Program: Dr. Phyllis Levenstein

Dr. Levenstein's program used commercially available toys taken by "Toy Demonstrators" to homes in an urban area of Long Island. Visits with mothers and infants on a weekly basis concentrated on improving verbal interaction, attempting to prevent educational disadvantage through early cognitive and affective intervention.

Experimental Variation of Head Start Curricula: Dr. Louise Miller

Dr. Miller's study compared the benefits of four established preschool programs: Montessori, Bereiter-Englemann, DARCEE, and a traditional nursery school program. Each group attended a six-hour daily program at age four, followed by either a traditional or academic kindergarten program at age five. One half of the DARCEE group received home visits in addition to the center-based program.

Harlem Training Project: Dr. Francis Palmer

Dr. Palmer's study provided one-to-one center-based intervention following two models: a Concept Training Group and a Discovery Group. The experimental children attended one or the other of the programs for one hour, twice weekly, for eight months.

Perry Preschool Project: Dr. David Weikart

This program provided academically high-risk children with a cognitively-oriented preschool program for two years before the children entered kindergarten. Five cohorts of children were studied over a period of thirteen years.

Curriculum Demonstration Project: Dr. David Weikart

This project utilized three curricula: Bereiter-Engelmann, cognitive training, and a unit-based or traditional model. The children attended a half-day program and were visited by a teacher once a week over a two-year period for 90-minute periods of instruction.

Carnegie Infant Program: Dr. David Weikart

Dr. Weikart's home-based infant program provided a series of lessons to very young infants and their parents. The curriculum emphasized the developmental theory of Piaget and stressed the facilitation of the growth of mothers as teachers.

Micro-Social Learning System: Dr. Myron Woolman

Dr. Woolman studied the effects of a preschool program utilizing an arrangement of modular units in which children worked through a preplanned series of activities. They received periodic reinforcement as they completed each objective in a sequence. The program design also included a life-simulator space in which the children applied their newly learned skills in free play. This aspect of the program utilized materials and equipment designed to provide unstructured free response favoring interactive play.

Head Start and Follow Through New Haven Study: Dr. Edward Zigler

This study investigated the effects of regular Head Start and Follow Through interventions on two cohorts of preschool children in New Haven, Connecticut, using measures of academic achievement, IQ, and social-emotional development. The original group has now been followed through the eighth grade.

Table 1: Summary of Studies

PROGRAM	PRINCIPAL INVESTIGATOR	LOCATION	TYPE OF DELIVERY SYSTEM	AGE AT INTERVENTION	YEARS OF PROGRAM
The Philadelphia Project	Dr. Kuno Beller	Philadelphia	center-based	4-6	early '60's
Institute for Developmental Studies	Dr.'s Martin & Cynthia Deutsch	Harlem	center-based	4-8	late '50's early '60's
The Parent Education Program	Dr. Ira Gordon	northern Florida	home-based	3 mo - 3	mid '60's
The Early Training Project	Dr. Susan Gray	Murfreesboro or Columbia, Tenn.	combination	4-5	early '60's
The Family-Oriented Home Visitor Program	Dr. Susan Gray	Nashville, Tenn.	home-based	0, 1	early '70's
Curriculum Comparison Study	Dr. Merle Karnes	Champaign - Urbana, Ill.	center-based	4	mid '60's
The Mother-Child Home Program	Dr. Phyllis Levenstein	Long Island	home-based	2-3	late '60's early '70's
Experimental Variation of Head Start Curricula	Dr. Louise Miller	Louisville, Ky.	center-based combination	4	mid 60's
Harlem Training Project	Dr. Frank Palmer	Harlem	center-based	2-3	mid 60's
Perry Preschool Project	Dr. David Weikart	Ypsilanti, Mich.	combination	3-4	early '60's
Curriculum Demonstration Project	Dr. David Weikart	Ypsilanti, Mich.	combination	4	mid 60's
Carnegie Infant Program	Dr. David Weikart	Ypsilanti, Mich.	home-based	3 mo - 2	late '60's
Micro-Social Learning System	Dr. Myron Woolman	Vineland, N.J.	center-based	4-5	late '60's
Head Start & Follow Through New Haven Study	Dr. Edward Zigler	New Haven, Conn.	center-based	5	mid '60's

D. Consortium Problems, Decisions, and Notes on the Environment of the Study

Once the basic decision to collaborate in a longitudinal study was undertaken, and it was agreed that all the original and follow-up data would be treated by a group who had no vested interest in the outcomes and had not participated in any of the original studies, three major kinds of decisions needed to be made:

- The variables and measures to be included in the follow-up,
- Decisions and choices in statistical treatment of the data,
- Reporting, dissemination and interpretation of the data.

The Decisions on Data to Be Collected.

The major constraint in the conduct of this study was fiscal. It was extremely difficult to find support for this research for a variety of reasons:

- There had been widespread publicity about the Westinghouse-Ohio University Study which was interpreted as meaning that there were no educational benefits from preschool education. The methodological problems in that study, built into the design prescribed by OEO were never widely understood. Private foundations we approached did not believe we could find any effects after ten years. The interpretations by Jensen, Jencks, Hernstein and Bronfenbrenner reenforced their expectation of negative findings.
- Many federal personnel were fearful that another follow-up study would simply serve to hammer the final nails into the coffin then being built for Head Start by the Administration in 1975.

- Many people believed that a consortium of this kind could never work - that a dozen senior and independent investigators would never agree on a single study design, or stick with it.
- Others felt that after a decade it would be impossible to find enough of the original samples to have any meaningful data.

There was interest among sophisticated decision makers in having something substantial upon which to base decisions about programs for young children.

- Governors were faced with decisions on state investments in services for young children.
- There was considerable pressure to make permanent the elimination of educational requirements from the Federal Interagency Day Care Requirements. This pressure was "justified" by the assertion that early education had no benefits worth the cost.
- Thoughtful politicians of both the right and the left questioned the assertion of no benefit, noting that half of middle class families purchase preschool services and have for almost a century. It seemed unlikely that so many consumers could be so wrong for so long a time.

The Office of Child Development agreed to take the risk of financing what many thought to be a hopeless task. The amount of money they could make available while substantial in toto, meant that each investigator would have to locate, interview, test, gain access to, and record school data on youngsters, and interview their parents for a total cost of about $250 per family. This is a small fraction of the cost-per-subject of other longitudinal studies. This limitation meant that we could not afford to make more than one contact per subject. That in turn limited the amount of information which could be collected.

Consortium members were torn between their personal wish to use that contact to examine socio-emotional variables and the public policy need to investigate cognitive variables. The public policy needs won out. Further, although IQ scores seemed not the most relevant of the cognitive measures which could be chosen, it was recognized that the prior public storms about IQ would cast doubt on the credibility of any other findings if IQ's were not also examined.

Thus the battery chosen: the WISC-R, the interviews, the school records, and, with reluctance, the school-administered achievement tests. Most of the group had grave doubts about our ability to use the achievement test data collected by the schools. Indeed the problem of treating those data has been the most vexing and time consuming methodological problem we have faced.

As it turned out, even the sparse allowance per-case we projected was wrong. We anticipated locating between a third and a half of the original subjects - if we were lucky. As the reader can see, we struck a jackpot and retrieved over 2/3 of the original subjects. And we kept running out of money.

Many of the institutions which employed the investigators, recognizing the unique and potentially important nature of the data, were persuaded to contribute all or most of their normal overhead costs. Several investigators gave up their summer salaries. Supplies, equipment, and telephone costs somehow got largely absorbed. Some of the deans and business offices developed sudden blind spots when passing the offices and research spaces co-opted by the consortium.

Indeed, as it turned out, keeping peace between the investigators was never a problem. Long standing differences were set aside, minimal time was spent on trivial debate, and decisions, once reached, were accepted. The Chairman's job turned out to be a combination of stringent fiscal management, begging, coping with school district lawyers and running interference with university administrators. A dedicated central staff coped with the data and kept in continuous touch with the research teams in the field.

Decisions and Choices in Statistical Treatment

Both secondary analyses and longitudinal research are activities noted for the size and frequency of their hazards and mazes. In combining both of them we knew that our own statistical skills would need supplementation. Many of the investigators, in addition to being leaders in Child Development, are superb methodologists and have excellent staff. All considered the problems their own and gave freely of their time and skill. We want to particularly acknowledge the assistance of John Madden (of Phyllis Levenstein's staff), Frank Palmer, and Mike Woolman.

Additionally we drew upon consultation from methodologists totally removed from the projects. Profs. Richard Darlington, Henry Ricciuti, John Doris, Steven Caldwell and Jay Millman, all of Cornell, were consistently generous with advice and information. Dr. Robert McCall, of the Boys Town Institute and formerly of The Fels Research Institute, shared his unique experiences with the treatment of longitudinal data. Dr. Bernard Brown, of OCD's research staff, was a constructive partner in every stage of this study, and helped us in innumerable ways. Prof. Urie Bronfenbrenner (also at Cornell) served as "heavy" critic, and periodically helped us re-examine the validity and meaning of our procedures. And finally, the Social Science Research Council's Committee on Longitudinal Methodology offered to critique the whole study.

Initially, believing that readers would understand that this was, of necessity, a heuristic investigation, we intended to use the usual methods of standard-score conversion to pool all of the data across studies. We rapidly discovered that finding positive effects touched a number of raw ideological nerves, and that we could expect to be attacked for not meeting the standards of a single, planned experiment carried out entirely in a laboratory with caged animals bred for the purpose of the study. While no field-based study of real people can meet the design criteria of a laboratory experiment, we came to the conclusion, after seeing the initial data, that in order to at least reduce resistance to these findings we would use the most rigorous and conservative data for treatment choices available. These choices have minimized the findings, just as the two sampling anomalies we found (i.e., Palmer and Miller) minimized differences between experimental and control populations.

Since there are virtually no other examples of so many fully independent tests of a single central hypothesis, conducted over so long a period in contemporary social science and brought together in terms of comparable data, we trust that the reasonably objective reader will understand what these findings are - and what they are not.

In general, Consortium members discussed treatment choices thoroughly, and helped central staff understand the limitations of their own data. Acting the way scientists are supposed to, they were remarkably non-defensive about their data, cooperated fully in providing information as best they could, and never attempted to interfere with the independence of the Cornell group in treating and interpreting the data.

Reporting, Disseminating, and Interpreting the Findings

At the first meeting of the Consortium, it was recognized that prompt reporting of findings was essential if they were to be useful in decision making. Work on two literature review documents was delegated immediately. Dr. Kuno Beller is chairing a subgroup preparing a technical monograph reviewing all available research on early intervention, which is awaiting these data for completion. Dr. Frank Palmer prepared a popular review of the early findings which was widely distributed in 1976. In June of 1977 he prepared a new review, presently being edited for publication, for the White House Commission on Mental Health. Dr. Bernard Brown successfully arranged for presentations by Consortium members and staff at major professional meetings across the country in 1976 and 1977 (See Appendix A-3). Over a thousand copies of a preliminary, non-technical report prepared in May, 1977 for an OCD meeting in El Paso, Texas have been distributed in response to requests from state and federal agencies, school districts, Head Start sponsors and universities across the country.

Substantial coverage by newspapers and newsletters has already occurred, and the Consortium chairman and staff have made presentations of the data to staff at the White House, in the Congress, and to various Head Start and day care groups. Articles have been scheduled in several major publications, and a summary of this report will appear in the widely read publication of The American Educational Research Association, The Educational Researcher.

If support for further multivariate analyses becomes available, dissemination of those findings will be underwritten by a grant from the Hewlett Foundation.

Additionally, the individual investigators will be publishing findings unique to their own studies, using data beyond the common core gathered by all the Consortium members.

A "popular" summary of this report is being prepared for distribution to state legislators and governors, and a slide/sound presentation of the data will be produced for loan to local groups.

A series of supplemental studies on special topics is planned to reach specialized audiences this fall and winter. Several doctoral dissertations are being drawn from these data, and will lead to additional publications. Consortium members have allowed themselves to be drafted for public and professional presentations, and a frequent use of telephones, air-package delivery, and finding ways to get together has kept the members reasonably abreast of each other's work.

Just as the data of each study remains the "property" of its principal investigator, so too has the independence of each investigator to publish and interpret the data in his/her own way been carefully preserved. Interpretations in this document are those of the Chairman and the staff, although the membership will be asked to review the document before its delivery to our sponsors.

The Consortium worked, we think, because its membership were personally secure, mature and well established scientists who respect each other, and were willing to forego personal goals in a genuine effort to produce a socially as well as scientifically important study.

Under such circumstances, decision making and implementation flowed from the importance of the task. Open knowledge of finances, difficulties, and limitations eliminated most problems of misunderstanding. The few possible sources of conflict were quickly brought to group discussion, despite the discomfort of such discussions. We believe that these practices, plus the character of the investigators, made this project possible.

E. Problems of Secondary Analysis

This report essentially constitutes a secondary analysis of the data from the Consortium members' projects. This is true even though the 1976-77 data were collected specifically for this report, since the design of the projects had been determined prior to the collection of these data.

The most basic problem of secondary analysis in general is that the analysis uses data in a manner for which they were not originally intended. A prime example of this occurrence is the importance of the Stanford-Binet scores in the current analysis. The Binet was used by the original projects as one of many measures of program effectiveness. However, since the Binet was the test used by more projects than any other (in fact, other than the Peabody Picture Vocabulary Test it was the only test used by a majority of projects), it assumes a much greater importance in the current analysis that it had in the original analyses by the projects.

A second problem of secondary research is that the questions which can be posed are limited by the data which have already been collected. One simply cannot ask whether early education affects second grade IQ scores unless the projects gave IQ tests at that time. This problem is made more complex by the fact that the projects were conducted independently and, hence, the chances are slim that most projects would have independently collected the same measure at the same time (in fact, this happened only for the Binet and Peabody Picture Vocabulary Test.)

A third major problem is that the experimental designs were different (particularly with respect to selection and assignment of program and control children) and that, hence, comparisons of the studies are always tenuous. A secondary analysis does not give as directly useful information as would a carefully designed experiment of the same size. (An analysis of the sample selection and assignment procedures used by each project is contained in Appendix B-1.)

F. Problems of Longitudinal Analysis

Since this report follows the same subjects over a number of years, it is also subject to the problems involved in longitudinal research.

The most basic problem in longitudinal research is that of attrition, i.e., the loss of subjects over time due to death, inability to be located, refusal to participate any longer, etc. This loss of subjects is only a minor problem if the loss is random, i.e., if there is no factor which causes the dropouts to be systematically different from the final sample. However, the results of research can be biased if one of three types of attrition occur:

- the dropouts differ from the final sample on some important characteristic (e.g., if children with less educated parents tended to drop out while those with more educated parents tended to stay in the final sample).

- differential attrition between groups - if the pattern of attrition differed between the program and control groups (e.g., if the children with less educated parents were those who dropped out of the control group, while the children with more educated parents tended to drop out of the program group).
- differential attrition due to refusals to participate by unhappy or dissatisfied subjects.

The overall evidence in our study seems to indicate that, although there are some instances of differential attrition, the final samples are representative of the original samples. Overall there were no significant differences between the demographic characteristics of those subjects found and those not found. The only significant attrition depended on whether specific pieces of data were in hand on July 1, 1977. Thus, attrition appears to have been a fairly random process. This conclusion is not surprising given the fact that much of the attrition occurred while neither program nor control children were receiving any major benefits (or costs) from the project (i.e., after project termination). The major reason to suspect biases caused by attrition is that the benefits received from the program will cause a different pattern of attrition in the program versus control groups. Since the current follow-up occurred several years after program termination, there is little reason to suspect that differential patterns of attrition would occur when neither group was receiving benefits from the program. The most crucial period for which attrition must be investigated in evaluation research is the period of program duration - and the question of attrition during the program has been addressed by most of the projects in their original reports which deal with more immediate posttests. The fact that less than 3% (N=22) of the parents and/or youths contacted refused to participate in the follow-up eliminates any concerns we had about contemporaneous differential attrition. The detailed attrition analysis is contained in Appendix B-3.

A second problem in longitudinal research is that the measures used can change in meaning over time. This is especially true of IQ and achievement tests since both the content and the standardization change with age. The content of a test used to measure the IQ of a seventeen year old obviously differs from the content of a test used to measure the IQ of a three year old. Given such differences, it takes a tremendous leap of faith in the infallibility of the test designers to say that an IQ of 100 at age seventeen is "the same as" an IQ of 100 at age three. This problem is further complicated by the fact that IQ and achievement tests are standardized cross-sectionally (e.g., IQ at age three is standardized on a group of three year olds, IQ at age four on a different group of four year olds, etc.) rather than longitudinally (e.g., IQ at age three being standardized on a group of three year olds, IQ at age four on the same group when they turn four, etc.). Cross-sectional standardization requires the additional leap of faith that the group of three year olds is "the same as" the group of four year olds. To avoid these difficulties, this analysis always considers IQ scores of the program children as compared only with the IQ scores of a control group. The posttest of the program group is never compared to their pretest as an indicator of whether they gained in IQ. Rather, the program children are said to have gained in IQ only if they have gained relative to their control group.*

A third problem in longitudinal analysis is that of test-retest effects, i.e., the fact that the very act of taking a first test may affect the scores of the second test. Again, this effect is controlled for to a certain extent in this report by comparing the program children only to their controls and not comparing program posttest with program pretest as an indicator of success. (This method controls only in those instances in which the control children were also given pretests; however, this did occur in most projects.)

G. Overview of the Analysis

Data for this study were collected in two stages. First, the original data collected independently by the projects were duplicated and sent to the central office. This original data consisted of any IQ, achievement or other psychological test data plus such demographic data as were collected by the projects. Second, each project collected current-year follow-up data in 1977 consisting of a standard parent interview, youth interview, school record data, achievement test data and WISC-R IQ data. The interviews and data collection forms were developed by the Consortium and were utilized by all but one project in this follow-up.

The analyses contained in this report are based on the Stanford-Binet IQ test scores (from the original data), the WISC-R IQ scores.

---

* We recognize that this is a more conservative standard than is considered necessary by many psychometricians. Throughout this study we have selected the more conservative of alternative data treatment choices, so as to avoid debates which are not germane to the principal problems being investigated.

indicators of assignment to Special Education and of grade failure from the school records, plus selected variables from the youth and parent interviews (from the current follow-up data). The status of the other data collected in this project is contained in Appendix A-4. Data collected and not reported in this paper either arrived too late for inclusion and will be included in later analyses, or have not yet been analyzed.

PART II

DATA COLLECTION METHODOLOGY

II. Data Collection Methodology

Central data collection methodology is described in this section. More specific information on the follow up data collection methodology (response rates, interviewer characteristics, etc.) is contained in Appendix B-2. The methods used in the various analyses are contained in the analysis sections.

A. Original Data

The "original data" consists of data on individual children collected by the projects prior to their becoming involved in the Consortium. Raw test and demographic data were duplicated at each project and transferred to the Central Office. Five projects (Beller, Levenstein, Miller, Weikart-Carnegie and Zigler) transferred all data on Hollerith cards. Six projects (Deutsch, Gordon, Karnes, Palmer, Weikart-Perry and Weikart Curriculum Demonstration) transferred most data on Hollerith cards plus some data (generally demographic information never before coded) on raw data sheets. Two projects (Gray ETP and Gray FOHV) transmitted all data on raw data sheets. Data from the raw data sheets were coded and keypunched at Cornell. All data were then checked for accuracy and internal consistency and, where necessary, corrected after consultation with personnel at the project from which they came. Crosstabulations of type of data by project were performed to select those variables which were collected by enough projects to be selected for common analyses. Based on the results of these analyses, the Stanford-Binet IQ scores plus selected demographic data were transformed to a common format and placed on a single SPSS file. In addition, separate SPSS files for each project containing essentially all data received were created.

For purposes of these analyses a case was defined as any subject whose test score was transmitted to the central office. In some instances, this definition varies slightly from the definition of a case used by the original projects, producing minor differences in the reported N's.

B. Follow-up Data

The decisions on the choice of common instruments to be used in the present follow-up study evolved from lengthy discussion of alternatives by the Consortium members which lasted for over six months. Because of financial restrictions and the high cost of locating subjects who had not been interviewed or tested for many years, it was necessary to choose a limited battery of measurements. The Consortium agreed to administer the age-appropriate Wechsler Intelligence Scale and the Consortium-developed Parent and Youth Interviews, and to obtain data from the achievement tests administered by the schools as well as data from the school records. (See Appendix B-6, "Instruments".)

1. Parent and Youth Interviews

The Developmental Continuity Parent Interview was initially developed from a compilation of interviews used by individual investigators, especially those of David Weikart and Martin Deutsch. The interview was pretested twice; the second pretest used a carefully controlled design which included families of children who had attended Head Start in Ithaca, New York, or had participated in a home-based parent education program based in Norwich, New York, from two to ten years earlier. The pretest sample parents were representative of the actual sample to be interviewed on such variables as race, urban/rural status, and number of years since their child had been in the program.*

The Parent Interview was designed to obtain comprehensive information on household composition, socio-economic status, parental attitudes toward, aspirations for, and evaluations of their child, information on the child's medical history, school educational history, the parent's current relationship with the child, and parental assessment of the intervention program.

The Youth Interview also drew upon previously developed interviews used by the principal investigators, but it leaned more heavily upon items of salient interest to participants at several Consortium meetings. This interview was also revised and pretested in the manner of the Parent Interview.

The Youth Interview obtains information on the child's status in school, his educational and occupational aspirations, leisure time activities and interests, employment status, and integration into his peer group and the larger community.

The interviewing, which had begun in October, 1976 for some sites, was completed by most sites before July 1, 1977. Two sites which were delaved for a variety of reasons (principally the problems associated with getting school record data) will complete their interviewing and testing this summer. As of August 15, 1977, 958 Parent and 910 Youth Interviews had been completed and sent to Cornell. It is expected that the total number of each type will exceed 1,200.

---

* None of the studies took place in upstate New York, and these sites were not included in this study except for pretests of the interview forms. Pretest interviewers and subjects varied in ethnic group membership.

All project sites experienced some difficulties locating their subjects after so many years. Various methods were used. Some projects began with name and address lists, both old and current, provided by local schools and Boards of Education (with elaborate procedures to insure informed consent by parents prior to the release of information). Other sites had maintained contact over the years through home visits, Christmas cards and post office forwarding addresses. Recently retired school teachers who lived in the community were a valuable source of the families' current addresses. Most sites contacted relations or friends of the families or asked the located parents about other families who were difficult to contact.

2. Wechsler Intelligence Scales

The Wechsler scores (WISC, WISC-R and WAIS) were sent to Cornell in a common format on coding sheets. A total of 725 Wechsler scores were received from seven project sites. Five projects (Gray, Karnes, Miller, Palmer and Woolman) administered WISC-R's during 1977 and sent them to Cornell for the final report. The Palmer project sent additional WISC-R scores which had been collected during 1976, and the WISC-R's sent for Gray's ETP group were administered in 1975. The Levenstein project sent WISC scores collected during the current year, while the Perry project sent WISC scores collected when the subjects were in eighth grade. (Since the Perry project consisted of five cohorts of children, it was decided that it was best to send test scores from a point when all children were the same age rather than from the current year.) In the remaining sites, WISC testing was not completed in time for inclusion in this report.

3. School Record Forms

For the October NERA meeting, Cornell staff developed an instrument for collecting school characteristics and student performance (marks, attendance, discipline). After Consortium members expressed concern about the time and money involved in collecting this data for every grade, the decision was made to develop a short form to be used by all sites and an optional long form. The School Record Form (face sheets, short form) provided data on the most critical variables: current status in school, placement in special education, retention history, or accelerated experiences. In addition, school identification by code number, location and type was collected so that school variables such as funding level, staffing, and ethnicity, for example, could be investigated through secondary sources at a later date. Each site agreed to provide school and achievement test data for four years (the latest available year and three other specified years) rather than every year. Information about the School Record Supplement is contained in Appendix B-2.

Most of the data on assignment to special education and on failure in grade were transferred by the staff of the individual investigators from the school records onto the School Record Form designed for use by all the projects. In some cases, however, the investigators had already collected such data from the school records in a different format, and these data were used in the analysis.

The sources of the data used in the special education and grade retention analyses are listed below.

- For Gordon's project, the analyzed data on special education and retention contained in the report "School Performance as a Function of Early Stimulation" were used.
- For Gray's, Karnes', Miller's, and Woolman's projects, the raw data on special education and retention from the School Record Form were used.
- For Levenstein's project, the raw data on special education and retention collected from school teachers' reports on the children, were used.
- For Palmer's project, the analyzed data on retention from the paper "The Effects of Early Childhood Intervention" were used.
- For Weikart's project, raw data from fourth grade collected on their forms were transcribed to the School Record Form.
- For Zigler's project, the raw data on retention collected in a different format were used.

Since some of the projects collected data on other than the School Record Form, the definition of "special education" varies somewhat across projects. The School Record Form itself provided four categories which are considered as special education for this analysis: special education (unspecified); educable mentally retarded or trainable mentally retarded; emotionally disturbed; and learning disabled classes. The definitions used, by project, are as follows:*

- Gordon: Educable mentally retarded, trainable mentally retarded, Specific learning disabled, emotionally disturbed
- Gray: Educable mentally retarded, trainable mentally retarded

---

* Speech and hearing classes were not included in the definition of special education since many of the problems which could cause assignment to such classes are not susceptible to remediation by education projects.

- Karnes: Special education (undefined), educable mentally retarded, Trainable mentally retarded, learning disabled, emotionally disturbed
- Levenstein: Educable mentally retarded, BOCES, learning disabled, remedial reading, remedial math, bilingual*
- Miller: Special education (undefined), educable mentally retarded, trainable mentally retarded, learning disabled, emotionally disturbed
- Woolman: Special education (undefined), educable mentally retarded, trainable mentally retarded, learning disabled, emotionally disturbed
- Weikart: Special education (undefined)

---

* A few children in bilingual classes were listed as being in special classes.

PART III

COMPARISON OF SAMPLES ACROSS PROJECTS

All projects in this study claimed to serve "disadvantaged" children. However, an obvious question which arises is this: to what extent does "disadvantaged" mean the same thing across projects? More precisely, were the children in one project comparable to the children in other projects? This question can be approached in two ways. First, the children in the projects can be compared on various measures to see if there are significant differences across projects. Secondly, the procedures and criteria for sample selection of each project can be analyzed. The first approach is dealt with in this section; the second in Appendix B-1.

The children in the different projects were compared on three measures (all collected at the onset of the projects): pretest Stanford-Binet I.Q., mother's education, and head of household SES (Hollingshead ISP). One way ANOVA was performed for each of the three variables, using project as the grouping variable. Each ANOVA was performed separtely on two samples: the original sample (i.e., at the beginning of the projects) and the current follow-up sample. Thus, a total of six ANOVA were run. Following ANOVA, the Scheffe procedure was used to divide the projects into homogeneous subsets (i.e., groups of projects among whom there were no significant differences).

Table 2 reports the results of the ANOVA's on pretest I.Q. for both the original and follow-up samples. In both cases there were significant differences among the projects. As seen in Figure 1, there were no significant differences for the original sample among the Miller, Karnes, Palmer, Levenstein, Deutsch, Beller or Gray projects. The Perry project was significantly lower in I.Q. than any other project except for CD, while the CD project was significantly lower than three of the other projects. Thus, the two projects which had specified a maximum I.Q. of 84 as a criterion for entrance (see Appendix B-1) had, as expected, the lowest IQ. As seen in Figure 2, the Perry project was significantly lower than all other projects for the follow-up sample and there were no significant differences among any of the other projects.

Table 3 reports the results of the ANOVA's on mother's education for both the original and follow-up samples. Again, the differences were significant. Figure 3 shows the groupings for the original sample, with Palmer's project having the highest mother's education (significantly higher than Deutsch, Karnes, Perry and ETP) and Perry and ETP having the lowest mother's education (significantly lower than Palmer, Zigler and Miller). Figure 4 shows the groupings for the follow-up data, with Palmer and Miller significantly higher than Weikart and Gray.

Table 4 shows significant differences in the ANOVA's in SES for both samples. Figure 5 shows that for the original samples, Palmer is significantly higher than six of the nine other projects, while Perry and ETP are significantly lower than Miller and Palmer. Otherwise the projects are homogeneous. Figure 6 shows that Palmer is significantly higher than all other projects in the follow-up samples and that Miller is significantly higher than Weikart and Gray.

In sum the analyses show that there are some significant differences among the projects on pretest IQ, mother's education and SES. However, the analyses also show that there are basically one or two projects in which the samples are significantly different than the rest, while the balance of the projects are homogeneous. The Perry and CD projects are lower in IQ than the others due to their maximum IQ criterion. The Palmer and Miller projects are generally the highest in mother's education and SES, while the Perry and ETP projects are lowest. Thus, in general, the children in the different projects are very similar on these measures. However, the differences which have been pointed out, particularly those of the Perry, Palmer, CD and ETP projects, should be kept in mind when evaluating the results of analyses.

Several words of caution should be added at this point. First, the measures used, of course, can give only a rough reflection of the true differences among the children. These measures do not reflect such crucial differences as the fact that, for instance, children in the ETP project grew up in Murfreesboro, Tennessee, in the early 60's, while children in the Palmer project grew up in Harlem in the late 60's. Although the ETP, Karnes, Perry and CD projects administered full Binet pretests to their entire samples, Levenstein administered a Binet pretest to only about 20% of her sample (others were given the Cattel) and Palmer administered a pretest to only 60% of his sample (those who began preschool at age two were not pretested). Miller did not administer "pretests" to her experimentals until they had already been in the programs for two to three months. Given the evidence that preschool significantly increases children's IQ's within the first three months, therefore, it is likely that the programs had already increased the children's IQ's by the time they were "pretested." Their true pretest scores would probably have been several points lower, and the "pretest" scores reported cannot validly be compared either with control pretests or with pretests for experimentals from other projects.* Furthermore, since neither Gordon, Gray-FOHV, nor the Carnegie projects gave Binet pretests (as their children entered before age three), these projects could not be included in this analysis.** Thus, the analysis of the differences in pretest I.Q. scores applies only to a limited number of projects.

---

* This "late" administration of the initial intelligence test may account for some of the differences in findings for Miller's groups as compared to the findings for the other studies.

** While it is possible to convert Binet and other tests to standard scores in order to include the other studies' pretest IQ's in this analysis, we have again chosen the most conservative option.

Table 2: Comparison of Projects on Pretest Stanford-Binet IQ

Project	Original Sample		Follow-up Sample	
	Mean I.Q.	N	Mean I.Q.	N
Beller	89.67	168	92.58	86
Deutsch	90.85	930	92.43	54
Gordon	-----	0	-----	0
Gray ETP	88.39	88	88.19	73
Gray FOHV	-----	0	-----	0
Karnes	92.30	224	94.73	122
Levenstein	91.18	45	92.05	21
Miller	92.37	247	93.17	127
Palmer	91.28	183	92.21	131
Weikart-Perry	79.02	123	79.02	123
Weikart-CD	80.56	41	-----	0
Weikart-Carnegie	-----	0	-----	0
Woolman	-----	0	-----	0
Zigler	-----	0	-----	0
average	90.12		90.25	
overall F	10.57		11.854	
significance	$<.01$		$<.01$	

Figure 1

Homogeneous Groups of Projects*
on Pretest Stanford-Binet IQ
for the Original Sample

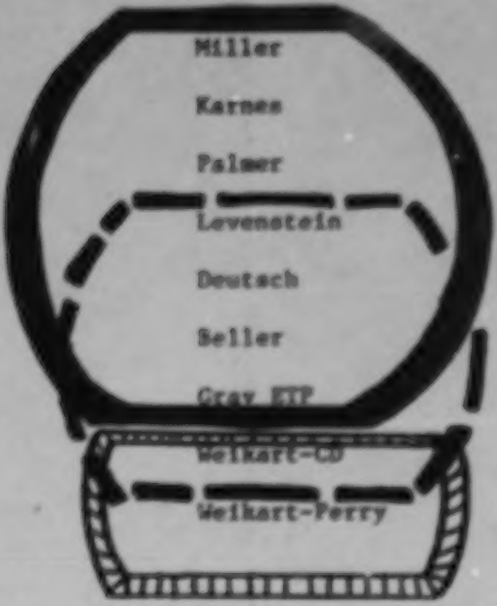
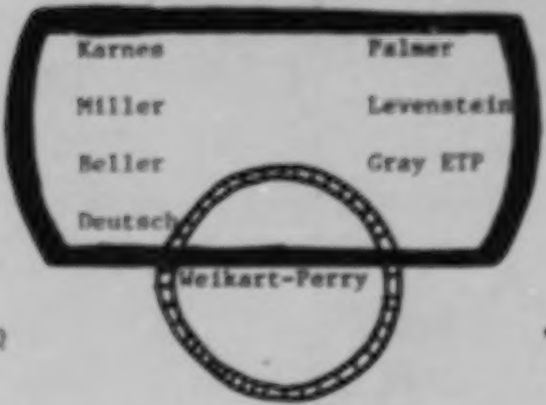

Miller
Karnes
Palmer
Levenstein
Deutsch
Beller
Gray ETP
Weikart-CD
Weikart-Perry

Figure 2

Homogeneous Groups of Projects*
on Pretest Stanford-Binet IQ
for the Follow-up Sample


Karnes
Palmer
Miller
Levenstein
Beller
Gray ETP
Deutsch
Weikart-Perry

▬ highest IQ group

▬ ▬ ▬ middle IQ group

▥ lowest IQ group

* within each circle (Venn diagram), there is no significant difference in IQ among the projects according to the Scheffé test

Table 3: Comparison of Projects on Mothers' Education

	Original Sample		Follow-up Sample	
	Mean Grades Completed	N	Mean Grades Completed	N
Beller	---	0	---	0
Deutsch	10.14	164	10.09	44
Gordon	10.49	115	10.68	37
Gray ETP	9.13	88	9.28	75
Gray FOHV	10.18	49	---	0
Karnes	10.11	200	10.08	118
Levenstein	10.51	248	10.53	185
Miller	10.82	227	10.79	117
Palmer	11.23	299	11.14	219
Weikart-Perry	9.42	123	9.42	123
Weikart-CD	10.34	41	---	0
Weikart-Carnegie	10.71	56	---	0
Woolman	---	0	---	0
Zigler	11.07	97	---	0
average	10.47		10.38	
overall F	11.431		6.995	
significance	$<.01$		$<.01$	

Figure 3

Homogeneous Groups of Projects*
on Mother's Education
for the Original Sample

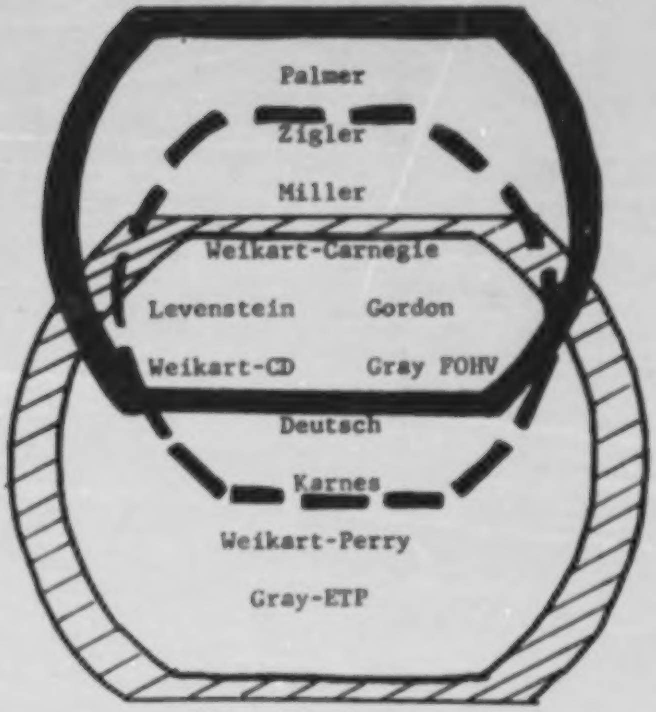
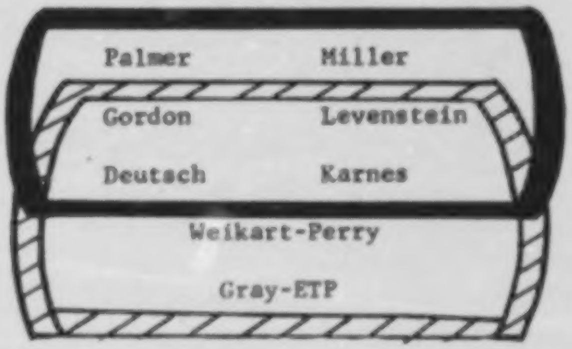

Palmer
Zigler
Miller
Weikart-Carnegie
Levenstein
Gordon
Weikart-CD
Gray POHV
Deutsch
Karnes
Weikart-Perry
Gray-ETP

Figure 4

Homogeneous Groups of Projects*
on Mother's Education
for the Follow-up Sample


Palmer
Miller
Gordon
Levenstein
Deutsch
Karnes
Weikart-Perry
Gray-ETP

▬ highest education group

▬ ▬ ▬ middle education group

▨ lowest education group

*Within each circle (Venn diagram), there is no significant difference in mother's education among the projects according to the Scheffe' test.

Table 4: Comparison of Projects on Head of Household SES

Project	Original Sample mean ISP	Original Sample N	Follow-up Sample mean ISP	Follow-up Sample N
Beller	---	0	---	0
Deutsch	65.94	163	64.88	43
Gordon	63.67	108	64.18	34
Gray ETP	69.68	82	69.67	70
Gray FOHV	66.41	46	---	0
Karnes	64.76	193	65.30	114
Levenstein	64.94	247	64.76	184
Miller	63.15	170	63.46	94
Palmer	58.17	295	59.18	216
Weikart-Perry	68.48	122	68.48	122
Weikart-CD	64.34	41	---	0
Weikart-Carnegie	---	0	---	0
Woolman	---	0	---	0
Zigler	---	0	---	0
average	63.95		64.21	
overall F	14.132		9.074	
significance	$< .01$		$< .01$	

Figure 5

Homogeneous Groups of Projects*
on Socio-economic Status
for the Original Sample

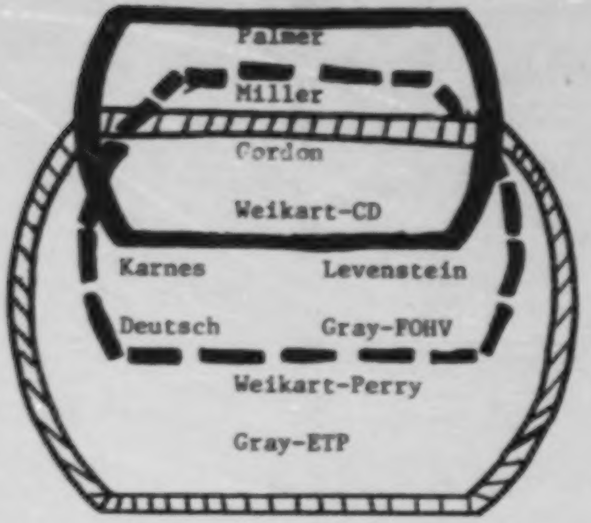
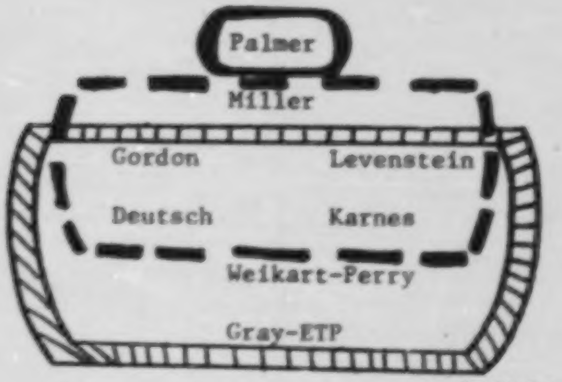

Palmer
Miller
Gordon
Weikart-CD
Karnes
Levenstein
Deutsch
Gray-FOHV
Weikart-Perry
Gray-ETP

Figure 6

Homogeneous Groups of Projects*
on Socio-economic Status
for the Follow-up Sample


Palmer
Miller
Gordon
Levenstein
Deutsch
Karnes
Weikart-Perry
Gray-ETP

■ highest SES group

▬ ▬ ▬ middle SES group

▥ lowest SES group

* Within each circle (Venn Diagram), there is no significant difference in SES among the projects according to the Scheffe′ test.

PART IV.

ANALYSIS OF PROGRAM AND CONTROL DIFFERENCES IN COGNITIVE FUNCTIONING

The first question one is inclined to ask when presented with the body of data collected by this project is this: overall, does the evidence indicate that early education improves the academic performance of low income children?* The fact that this report is a secondary analysis of independently conducted research efforts limits the rigor with which this (or any other) question can be addressed by this body of data. These data, however, contain better selected non-treated controls than any other study of comparable size. Additionally, the fact that it would take two decades to collect equivalent longitudinal data makes these data perhaps the best available to answer questions of the effectiveness of early education for low income children.

Due to the differences in sample assignment procedures, comparisons between program and control are made for each project individually; data are not pooled across projects in this report. Rather, in order to evaluate the overall evidence, the significance levels from the individual project program-control comparisons are pooled using a standard statistical method.

---

* All of the projects except the Perry Preschool and the Carnegie Infant Program compared variations of programs (either in terms of age at which the program was administered, or in terms of variations of the curriculum or approach, or in terms of both). However, this analysis deals primarily with the effect of participating in a preschool program in the context of family variables and does not make a systematic attempt to distinguish among types of preschool approaches. Therefore, these variations are grouped together as "having received experimental preschool education" and are contrasted only against controls who had not received such an education rather than against each other. It must be remembered that this secondary analysis has purposes different from the original intent of these studies.

Of the data accumulated in usable form by the time of this report, three bodies of data seemed particularly appropriate in answering the question of whether early education can produce academic benefits for low income children: the Stanford-Binet IQ data collected independently by most of the projects over a number of years; the WISC-R data collected primarily as part of the current follow-up; and the school record data on assignment to special education and grade retention collected primarily as part of the current follow-up. The analysis consists of three stages, with each stage corresponding to a particular body of data.

The first stage employs the Stanford-Binet IQ data to answer the question of whether the programs increased the mean IQ's of the children relative to a control group and, if so, for how long did these gains last? The question is addressed first by making simple comparisons (t-tests) of the mean posttest IQ scores of program and control groups for all available data points for each project. The significance levels for the comparisons are grouped by number of years after program termination and then pooled across projects. This technique gives an indication of whether, given all available evidence, there is a significant difference in IQ scores between program and control children a given number of years after the program. Secondly, the procedure is repeated while controlling for pretest IQ where available. This analysis serves primarily as a check on the sample assignment procedures of the projects.

The second stage consists of an analysis of the WISC-R IQ data collected during the current (or previous) year for those projects which have reported it thus far. This body of data constitutes the latest IQ data available and is useful primarily for addressing the question of whether early education can produce long-term IQ score increases. Since these data were collected at widely different ages (ranging from ten to seventeen years), the analyses are reported individually for each project. Again, the analyses consist of t-test comparisons of program and control group mean IQ's.

The third and final stage consists of the analysis of whether the children were assigned to special education classes or held back in grade. Crosstabulations of the number of children assigned to special education and of the number of children held back by program-control status are performed for each project and significance levels are pooled. This analysis indicates whether early education affects actual performance in school. The last section of Part IV discusses analysis of school related responses from the Youth Interview.

A. Analysis of Stanford-Binet IQ Scores from Immediate Posttest to Three Years after Program Termination

The Stanford-Binet was chosen because it was the test most commonly used by the Consortium projects and because it has had a fairly large scale standardization. In all projects the Binet was only one of many tests used to evaluate the program. The Binet evaluates a wide range of abilities, only a few of which may have been the targets of a given project. Thus, the project could have achieved a significant improvement in the abilities toward which it was aimed without having evidenced a significant change in the Binet score, since the Binet also taps other abilities which the project had not attempted to improve. A perusal of individual project reports reveals that in many instances the investigators found other tests to be more relevant than IQ to the purposes of their project. In fact these other tests often yielded more favorable results than the Binet. (Unfortunately, however, since such tests were more specific to the goals of each project, none of them was used by more than one or two projects; hence, they form a less solid basis for comparison of the projects.)

In using the Binet as a dependent variable, this paper is employing a very stringent criterion of success for these projects. Therefore a failure to improve the Binet IQ cannot be taken as a failure of a program to accomplish its goals. On the other hand, the success of a project in improving IQ scores means that it has accomplished at least one goal which our society deems important.

This analysis is also adopting a conservative tactic by comparing only Binet IQ's (rather than comparing results on Binet, WISC, Cattell and other IQ tests) due to the problems inherent in comparing one IQ test with another. This approach results in the loss of early scores in the Gordon, FOHV, VIP, and Carnegie projects, as well as several follow-up testing periods in the VIP project.* Problems of comparability across projects still occur because of differences in testing situations (e.g., differences due to sex and race of tester, or extent of efforts to let the child "adjust" to the testing situation). However, such differences are assumed to be negligible unless otherwise noted in the text.

IQ scores are reported on the basis of the 1960 deviation IQ tables because these were the scores reported by the individual projects.**

---

* Separate later analyses may be undertaken at a later date in which scores from various tests will be converted to standard scores and compared.

** Non-psychometrist readers should be cautious in comparing these 1960 IQ scores with Binet scores based on the 1972 standardization since they are not identical at the earliest age levels.

For the purpose of comparing testing periods across projects, one testing period "slot" was created for every year of age (e.g., the first slot ranged from eighteen to twenty-nine months, the midpoint being twenty-four months, etc.).* The testing periods were then assigned to the slots in which their mean chronological age fell. For example, any testing period with a mean chronological age between thirty and forty-one months would be compared with any other projects' testing periods where the mean chronological age was between thirty and forty-one months. This procedure ensures that only tests which were given at approximately the same ages are compared with each other and minimizes the age bias built into this test.**

The available Binet data allow one to evaluate the effects of preschool on IQ test scores up to three years after children have left the program.+ Table 5 gives the results of the simple t-test comparisons of mean posttest Stanford-Binet IQ scores for the program and control groups for each project. With only two exceptions (Miller at ages seven and eight), the mean IQ of the program groups is higher than the mean IQ of the control group. When the significance levels are grouped by length of time after the termination of the program* and are pooled using the technique recommended by Darlington** (as shown in Table 6), the evidence shows that early education can produce significant increases in IQ (over a control group) which last for up to three years after the child leaves the program. The pooled z scores are significant for the immediate posttest and for the one, two, and three years after posttests. Thus, the evidence from simple t-tests suggests that early education can indeed increase IQ scores and that these gains last for at least several years. One should note from Table 5, however, that for most projects, while the difference in IQ scores between program and control groups decreases over time, it does not decrease as immediately as had been previously asserted by critics of early intervention.

---

* Thus, all children in one project's testing period were always placed in the same slot. This procedure might seem too obvious to mention; however, specification of the procedure was necessary because some projects defined testing period by chronological age regardless of time in the school year, while others defined testing period by time of the school year (e.g., post first grade).

** An alternative procedure would have been to define testing period slots by number of years after the termination of the preschool; however, this would have entailed comparing IQ across different ages and this poses other problems.

+ Karnes and Miller have administered both the Binet and WISC-R to samples of the 1977 follow-up group. The new Binet data did not arrive in time for this report, but may be treated and reported separately later this year.

The next question which must be asked is this: does the effect of early intervention on IQ remain after one has controlled for the influence of pretest IQ scores? This question is essentially an evaluation of the sample assignment procedures. If the procedures functioned to insure an initial equivalency between groups in IQ scores, then the procedure of controlling for initial IQ will yield essentially the same results as would a simple t-test. To control for pretest IQ, the pretest Binet (or in the case of Levenstein, PPVT) IQ scores are introduced, along with program-control status, into a multiple regression equation determining posttest IQ. The equation used is as follows:*

---

* The significance levels were grouped by number of years after the termination of the program rather than by age for two reasons: first, if one pooled by age, it would entail pooling the results of immediate post-tests with those of later follow-up tests (which would lead to inaccurate inference due to the "fadeout" effect often observed in preschool programs for low income children) and second, the technique of pooling only significance levels circumvents the problem of comparing IQ tests across ages. Furthermore, analyzing by the number of years after program termination answers the question of how long the effects last as opposed to the question of to what age the effects last.

** Richard Darlington, Radicals and Squares, Logan Hill Press, Ithaca, NY, 1975, p. 525.

$$\text{IQ}_{\text{posttest}} = \text{PC} + \text{IQ}_{\text{pretest}} + (\text{IQ})^{2}_{\text{pretest}}$$

where PC = 1 for program children
PC = 0 for control children

The regression equation was run for each project for each posttest. The regression coefficient for the early education variable gives the difference in IQ points between program and control groups after controlling for the pretest. The significance of this coefficient tells whether the program had a significant effect on IQ after controlling for the effect of prior IQ. Hence, comparing the significance of the regression coefficient to the significance of the t-test shows whether controlling for initial IQ scores will alter the conclusions drawn from the t-tests about the effect of early education IQ.

This procedure cannot be carried out for all projects. Gordon, FOHV, and Carnegie did not give Stanford-Binet or PPVT pretests. Their children entered at too young an age for these tests and, hence, were given infant tests. Palmer pretested only half of his program group, and his control group was pretested at an average age of two years, eight months, while the program group was pretested at an average age of three years. Given the fact that two years of age is the very earliest the Binet can be given, plus the fact that the longitudinal standardization of IQ tests is questionable (see Methodology), the four months average difference could bias the pretest scores in a direction unfavorable to the hypothesis that the program would have long-term effects. As already mentioned (see Methodology), Miller did not administer "pretests" to her program children until they had already been in the program for two to three months. Thus, there appear to exist biases in both the Miller and Palmer projects which would bias results in favor of the controls when pretest IQ is entered into the equation.

---

* Both pretest IQ and pretest IQ squared are introduced into the equation in order to allow for a nonlinear relationship between pretest and posttest (i.e., to allow the line graphing the relationship to curve at some point) and thus provide a more unbiased estimator of the effect of preschool.

Table 7 compares the significance of the regression coefficients with the significance of the t-tests in all cases where pretests were reported. As can be seen, there was no change in significance in twenty out of the twenty-seven comparisons. In five cases, the regression coefficient was not significant where the t-test had been. However, two of these cases were from the Palmer project, for which the comparison of program and control, while controlling for pretest, may not be valid. Of the other two cases which showed a change in significance, one (Miller at age eight) was another case where controlling for pretest may not be valid, while the other (Levenstein at age five) actually showed a difference which became significant when the pretest was used as a controlling variable.

Thus, there appear to be only three of twenty-seven comparisons where controlling for the pretest eliminates the significance of the effect of early education on IQ. Moreover, in the two ETP comparisons, the regression coefficient was still quite close to the .05 significance level (.13 and .06).

In order to determine whether controlling for pretest IQ affects the pooled results that were reported in Table 6, the significance levels for the regression coefficients were, wherever available, substituted into the pooled z analysis. The results are shown in Table 8. (Table 8 thus contains the significance levels for regression coefficients where available and the significance levels for t-tests where the regression coefficients were unavailable). The results show that, although the pooled z's were somewhat reduced, there are still significant differences in IQ for up to three years. This is true even though the regression coefficients for the Miller and Palmer projects (which presumably contain biases favoring the control groups as explained above) were included in the analysis. While this procedure is not as rigorous as one would have designed in a primary analysis, it does indicate that early education does have a significant and lasting (for at least three years) effect on IQ. Other analyses conducted during this project (see Harry Murray, "Early Intervention in the Context of Family Characteristics," and Virginia Ruth Hubbell, "Differential Effects of Early Childhood Intervention Programs," Appendix C) show that this conclusion also holds when controlling for family environment variables. The overall evidence thus serves to refute the notion that early education has only a very short-term impact on the IQ of low income children.

Table 5: Mean IQ's for Experimental and Control by Project

	Age	Mean IQ Program	Mean IQ Control	t-value	Significance	N Program	N Control
Beller	Pre	92.16	88.39	1.93	.056	57	111
	5	98.59	88.39	4.48	.0001	56	111
	6	98.36	92.07	2.69	.008	53	110
	7	97.83	90.59	3.20	.002	52	105
	8	97.61	91.03	3.10	.002	51	99
	9	98.44	90.04	4.08	.0001	50	99
Deutsch	5		92.08	5.82	.0001	260	142
	6	100.24	93.14	4.03	.0001	189	97
	7	96.10	93.00	0.78	.437	31	21
	8	---	---	---	---	---	---
	9	97.63	93.92	0.88	.386	32	12
Gordon	3	94.88	91.40	1.82	.072	144	52
	4	93.89	88.47	2.31	.022	133	53
	5	94.02	88.94	2.38	.018	134	53
	6	94.01	88.69	2.72	.007	127	52
Gray ETP	Pre	89.43	87.34	0.74	.462	44	44
	5	96.05	86.32	3.96	.0001	47	43
	6	94.98	81.89	5.30	.0001	45	42
	7	97.71	89.56	3.07	.003	41	41
	8	93.63	86.10	2.47	.016	42	38
	10	88.42	81.17	2.63	.010	42	38
Gray FOHV	3	87.31	81.14	1.42	.162	29	14
	4	91.11	83.88	1.77	.083	35	16
	5	91.03	89.25	0.36	.724	32	16
Levenstein	3	101.14	92.72	3.74	.0001	124	29
	4	105.44	96.00	3.94	.0001	108	42
	5	107.07	103.17	0.74	.462	97	6
	6	101.44	87.30	2.12	.060	99	10
Miller	4	92.87	89.21	1.57	.117	213	34
	5	97.33	90.00	3.27	.001	210	34
	6	94.06	93.74	0.14	.891	200	31
	7	92.48	92.48	0.00	-.999	192	29
	8	89.72	94.00	-1.38	-.176	174	29
Palmer	3	91.63	85.32	3.05	.003	228	63
	4	97.74	93.31	2.17	.031	221	59
	5	96.41	91.13	2.55	.011	207	48
Weikart-Perry	Pre	79.57	78.54	0.89	.378	65	58
	5	94.05	83.20	5.40	.0001	64	57
	6	91.25	86.34	2.43	.017	64	56
	7	91.72	87.07	2.32	.022	61	58
	8	88.11	86.85	0.57	.569	62	56
	9	87.71	86.77	0.43	.666	61	56
	10	84.98	84.61	0.17	.861	57	57
Weikart-Carnegie	3	104.10	100.44	0.82	.417	20	18
	6	111.95	109.47	0.52	.606	19	17
	7	106.40	104.06	0.54	.594	20	33

Table 6: Comparing Program and Control on Stanford-Binet IQ Score - Pooled z's and Significance for t-tests by Number of Years after Termination of Early Education

	Immediate Posttest		One Year After		Two Years After		3 or 4 Years After	
	z score	Significance	z score	Significance	z score	Significance	z score	Significance*
Beller	4.48	.0001	2.69	.008	3.20	.002	4.08	.0001
Deutsch	1.69	.092						
Gordon	1.82	.072	2.31	.022	2.38	.018	2.72	.007
Gray ETP	5.30	.0001	2.96	.003	2.41	.016	2.58	.010
Gray FOHV	1.42	.162	1.73	.083	0.35	.724		
Levenstein	3.74	.0001	0.73	.462	1.88	.060		
Miller	3.27	.001	0.14	.891	-0.00	.999	-1.38	.176
Palmer	2.17	.031	2.55	.011				
Weikart-Perry	5.40	.0001	2.39	.017	2.29	.022	0.57	.569
Weikart-Carnegie			0.81	.417			0.52	.606
z =	9.76		5.44		4.73		3.71	
p =	.0000		.0000		.0000		.0002	

* All significance levels are two-tailed.

Table 7: Comparison of t-test Results with Regression on Effect of Program

	Age	Mean Difference	Regression Coefficient	t-value t-test	t-value Regression Coefficient	Significance t-test	Significance Preschool Regression
Beller	5	10.20	-----	4.48	-----	.0001	-----
	6	6.29	4.23	2.69	2.59	.008	.0096
	7	7.24	5.08	3.20	3.36	.002	.0008
	8	6.58	3.60	3.10	2.28	.002	.0226
	9	8.40	5.75	4.08	3.62	.0001	.0004
Deutsch	5	7.53	5.36	7.81	4.96	.0001	.0001
	6	10.89	3.53	10.36	2.07	.0001	.0384
	7	6.58	-----	3.74	NE	.0001	.99
	9	4.98	6.24	1.70	1.11	.092	.267
Gray ETP	5	9.73	7.85	3.96	4.37	.0001	.0001
	6	13.09	11.52	5.30	6.60	.0001	.0001
	7	8.15	6.29	3.07	2.80	.003	.0052
	8	7.53	3.91	2.[illegible]7	1.50	.016	.1336
	10	7.25	4.44	2.63	1.87	.010	.0614
Levenstein	3	8.42		3.74	-----	.0001	
	4	9.44	7.94	3.94	3.36	.0001	.0008
	5	3.90	3.89	0.74	0.82	.462	.4122
	6	14.14	12.08	2.12	2.84	.060	
Miller	5	7.33	4.81	3.27	3.05	.001	.0022
	6	0.32	-1.53	-0.14	-0.88	.891	-.3788
	7	0.00	-2.80	0.00	-1.40	.999	-.1616
	8	-4.28	-6.76	-1.38	-3.31	.176	-.001
Palmer	3	6.31				.003	
	4	4.43	-1.23		0.72	.031	.4716
	5	5.28	1.14	2.55	0.57	.011	.5686
Perry	5	10.85	10.47	5.40	5.54	.0001	.0001
	6	4.91	4.47	2.43	2.43	.017	.017
	7	4.65	4.05	2.32	2.16	.022	.031
	8	1.26	1.16	0.57	0.55	.569	.582
	9	0.94	0.70	0.43	0.35	.667	.726
	10	0.37	NE	0.17	NE	.865	NE

Table 8: Pooled Z's Substituting in Regression Coefficient Where Available

	Immediate Post		1 Year After		2 Years After		3 or 4 Years After	
	z score	Significance	z score	Significance	z score	Significance	z score	Significance
Beller	4.48	.0001	2.59	.0096]	3.35	.0008]	3.55	.0004]
Deutsch	1.11	.267]						
Gordon	1.82	.072	2.31	.022	2.38	.018	2.72	.007
Gray ETP	4.37	.0001]	2.80	.0052]	1.50	.1336]	1.87	.0614]
Gray FOHV	1.42	.162	1.73	.083	0.35	.724		
Levenstein	3.36	.0008]	0.82	.4122]	2.83	.0046]		
Miller	3.05	.0022]	-0.88	-.3788]	-1.40	-.1616]	-3.30	-.001]
Palmer	0.72	.472]	0.57	.569]				
Weikart-Perry	5.54	.0001]	2.39	.017]	2.16	.031]	0.55	.582]
Weikart-Carnegie			0.81	.417			0.52	.606
	8.62		4.38		4.22		2.41	

] = significance and z score taken from regression

All significance levels are two-tailed.

B. Analysis of WISC-R IQ Scores

As part of the current follow-up, WISC-R IQ tests were administered by the projects. Seven projects have thus far reported results. Of these, six will be treated here. (The seventh, Karnes, has no untreated control group). Of the six, one (Woolman) will be discussed separately due to the nature of his control groups.

Table 9 gives the t-test comparison of the mean IQ scores for the Levenstein, Palmer, Miller, Perry and ETP projects. Figures are given for the three types of IQ scores available from the WISC-R: Full Scale IQ, Verbal IQ, and Performance IQ.* As can be seen there are significant differences only for the two projects with younger children. The FIQ and VIQ are significantly greater for the program group in the project with the youngest children (Levenstein), while only the PIQ is significantly greater in the other (Palmer). For the projects with children twelve years, eight months or older there is no significant difference between program and controls (although the Miller results in this respect may again be a function of the selection process used for the control group). Thus, the results from these projects indicate that the effect of early education on IQ, although shown to be long term in the previous section, is probably not permanent. Analysis of the WISC-R subtest scores does nothing to alter this conclusion. Only in the Levenstein project were the program children significantly higher then the controls on more than one subtest. It should be noted, however, in no instances were the controls significantly higher than the program children.*

---

* Each test is made up of ten to twelve subtests grouped into a verbal scale (VIQ) and a performance scale (PIQ), each with five to six subtests. The full scale IQ (FIQ) is the sum of all subtests.

The Woolman project compared the program children to a random sample of other children in their classrooms. Since the program children were originally selected by the school district as disadvantaged, this comparison group would presumably be more advantaged. (The city of Vineland is not overwhelmingly poor. The 1970 Census reports 8.2% as being below the low income level.) As shown in Table 9, there is no significant difference in IQ between the program children and the other children in their classrooms. Thus, the children, although selected as the most disadvantaged by the school district, are not lower in IQ than the other children in the classroom. The children are an average of eleven years and three months old, so these results also support the hypothesis of IQ increases which last for several years.

---

* Examination of subtest score patterns may be undertaken later this year.

Table 9: Comparison of WISC IQ's by Project

Project		Mean Age	Mean IQ Program	Mean IQ Control	t Value	Significance	N Program	N Control
Levenstein	FIQ	9-9	101.86	93.56	3.21	.002	51	25
	VIQ	9-9	98.41	89.36	3.66	.0001	51	25
	PIQ	9-9	105.45	99.48	1.89	.063	51	25
Palmer	FIQ	12-2	92.31	87.64	1.65	.102	94	25
	VIQ	12-2	93.33	92.20	0.37	.711	94	25
	PIQ	12-2	92.71	85.20	2.53	.013	94	25
Miller	FIQ	12-8	84.96	87.69	-1.13	-.262	109	32
	VIQ	12-8	83.07	85.53	-0.96	-.337	109	32
	PIQ	12-8	89.42	92.41	-1.12	-.263	109	32
Perry	FIQ	14-0	81.02	80.71	0.14	.885	54	56
	VIQ	14-0	78.33	77.64	0.36	.721	54	56
	PIQ	14-0	87.59	87.82	-0.10	-.924	54	56
Gray ETP	FIQ	16-9	78.74	77.79	0.32	.752	34	38
	VIQ	16-9	77.09	78.47	-0.49	-.624	34	38
	PIQ	16-9	83.41	80.26	1.02	.310	34	38
Woolman*	FIQ	11-3	91.15	93.50	-0.76	-.451	34	40
	VIQ	11-3	86.82	91.33	-1.31	-.193	34	40
	PIQ	11-3	97.85	97.38	0.16	.873	34	40

* The control group measure is the mean IQ of the other children in the classrooms.

C. Summary of Results of the IQ Testing

Taken as a whole, the Stanford-Binet results from the original data and the WISC-R results from the current follow-up data refute two hypotheses:

(a) the hypothesis that early education has only a short-lived effect on IQ scores (i.e., that such programs basically train children to take tests and that the effects fade quickly as soon as the children are removed from the program).

(b) the hypothesis that early education has a permanent effect on IQ scores. The evidence seems to indicate that, on the average, well-run early education programs can increase a child's ability to perform well on IQ tests and that this increase in skills lasts for several years but eventually fades. The effect of the early education then, in terms of skills which are measurable by IQ tests, is certainly of far more value than a temporary fluctuation in skill levels but does not constitute a permanent increase in intellectual skills. The children's ability to deal with intellectual problems, then, is genuinely increased for a number of years. The child is not simply geared up to take a particular test.

The obvious question which arises from this evidence is whether such an increase in skills is valuable. This question is not directly answerable in this report; however, a related question can presumably shed some light on it. Specifically, did these programs enable children to function better in school than the control children? This effect could have been the product of the effect on IQ or of other effects of the programs (such as social effects or effects on parental attitudes). The following section will attempt to answer this question.

D. Analysis of Assignment to Special Education and of Grade Failure

Intelligence tests are designed to predict academic achievement. Obviously actual achievement is a better measure than is a predictive instrument. In this section we examine two indicators of actual school performance.

This section investigates whether low-income children who attend early education programs are less likely to be assigned to speical education classes, and whether they are less likely to be held back in grade. We found that early education does improve the ability of low-income children to meet the minimal requirements of their schools. The analysis is based on the results of eight independently conducted early education programs which have examined the school records of the children who attended the programs and of control children.

The use of these outcomes (whether a child was held back in grade and assignment to special education classes) as measures of the effectiveness of early education programs has one major advantage over the use of IQ or achievement tests in that grade failure and special education placement are concrete indicators of whether a child has performed acceptably within his/her educational institution.

The use of these outcome measures, however, does have a drawback in that both outcomes are affected by the policies of the individual states and school districts. While the subjects were originally living in a very small number of communities, by the time of this follow-up they had attended several hundred different schools. We do not have information on the distribution of promotion policies or the relative availability of special education classes in most of these school districts. We are assuming that with so large a number they are representative of the national situation, about which some information is available. Moreover, since both special education and grade failure are methods for dealing with children who do not meet minimal school requirements, the two may overlap to some extent as outcomes. For instance, if one project finds a significant effect on assignment to special education, but no significant effect on being held back in grade, while a second project finds the reverse, one could attribute these results to the fact that the former school district tends to place low achievement children in special education classes while the latter district tends to retain them in grade for a year. In fact, of the projects which have reported both special education and grade failure data thus far, none have had significant effects on both outcomes; however, all projects but one have had effects with a significance level of .10 or less on one of the two outcomes.

The combined evidence from five projects which looked at whether the children had been assigned to special education classes shows that early education significantly reduces the number of children assigned to special education.

The combined evidence from seven projects which looked at whether the children had been held back in grade indicates that early education significantly reduces the number of children held back.

An additional project, which compared the high-risk children who attended its program to the general school population, found that the program children were held back no more often than the general school population. Further, there was no significant difference in the rate of assignment to special education between program children and the general school population.

The importance of reducing the number of children assigned to special education or retained in grade is so great that it scarcely seems necessary to elucidate it here. The value of reducing the number of children who must undergo the trauma of being labeled "educable mentally retarded" or of failing a grade - as well as the value of the cost savings to the school systems - is self-evident.

The same analytic technique was applied to both special education and grade failure data. Crosstabulations of whether the child was in the program with outcome (e.g., whether the child was assigned to special education) are performed for each project individually. Yates' adjusted chi-square statistic is computed. The significance levels are then pooled across projects, giving an indication of whether the overall evidence when considering all projects together (i.e., considering each project as an independent replication of the others) shows a significant difference. The pooling of significance levels is accomplished in two stages. First, significance levels are pooled only for those projects which had designs which were very near to being experimental. Second, significance levels are pooled for all projects, including those which were more quasi-experimental.

These analyses are organized in four sections:

1. special education results of five studies
2. grade failure results of seven studies
3. results for the Micro-Social Learning System
4. summary and implications

1. Special Education

Data from five projects comparing the incidence of assignment to special education between program children and control children presents strong evidence that preschool education for low-income children reduces the number of children assigned to special education during their school years.

For each of the five projects, a crosstabulation of whether the child was in the program and whether (s)he was assigned to special education was performed and the Yates adjusted chi-square was computed. Results are shown in Table 10. The Gordon[*], Gray and Levenstein projects all show significant reductions in the number of children assigned to special education. The Perry Preschool project shows a two-tailed significance level of .096 and, thus, comes very close to being statistically significant (i.e., having a significance level of .05 or less). For all four projects, the differences in the percent of children assigned to special education are striking: for Gordon, 9% of program vs. 30% of controls; for Gray, 3% of program vs. 29% of controls; for Levenstein, 10% of program vs. 39% of controls; and for Weikart, 14% of program vs. 28% of controls. Only the Miller project gives no indication of a reduction in the number of children assigned to special education.[**] The pooled significance level for the five projects (using the technique recommended by Darlington) is .0002. Thus, the overall evidence of the five projects (comparing 461 children) shows that early education does significantly reduce the number of children in special education.

---

[*] The two-tailed significance level for the Gordon project was .052. One-tailed significance tests result in significant differences for all but the Miller project.

[**] See note on Miller data in Appendix B-1.

It is important to note that the projects with the most nearly random procedures for assigning children to program or control groups had, in general, the most impressive results. The Gordon, Gray and Weikart projects had the most stringent sample assignment procedures (i.e., most closely approximating an experimental design). Pooling only the results of these projects gives a significance level of .0006, again indicating a statistically significant reduction. Thus, it seems fair to say that the results with respect to special education are not due to any self-selection factor. The only project in which a difference was not found (Miller's) had a nonexperimental procedure for creating a control group, due to the fact that this project was primarily concerned with comparing different curricula, not comparing program children with those who had not received any program.* (See Appendix B-1 for summaries of the sample assignment procedures of each project.)

Perhaps the most interesting column in Table 10 is the last column (graphed in Figure 8), which gives the percentage reduction in children in special education for each project (e.g., the Gordon data estimates that 69% of the low-income children who could normally be expected to be in special education would not need to be in special education if they were given early stimulation). Although the number of children in a given study is small, in the data from all projects (except Miller's) the percentage reduction is very substantial (from 50.2% to 90.5%). Four of the projects reduced the expected number of children in special education by more than 50%. It is unlikely that these findings would be affected by sample fluctuations.

---

* According to a recent report by the Jefferson County (Louisville) School Board based on the 1976-1977 school year, the overall district rate of assignment of blacks to special education classes was higher than that of Miller's experimental groups.

In addition to the above results, the data from the Miller and Karnes projects, which were designed to compare different preschool curricula, indicate no significant differences in the rate of assignment to special education related to different curricula (see Table 11). Interestingly enough, the two projects showed similar results for the three curricula compared by both projects; of the three, Bereiter-Engelmann had the greatest percentage assigned to special education, Montessori was in the middle, and traditional had the smallest percentage assigned.* These results, though consistent, were not statistically significant for either project.

2. Grade Failure

Data from seven projects comparing the percentage of children held back in grade once or more between program and control groups presents moderate evidence that early education can have an effect on whether or not children are retained in grade.

---

* Miller's DARCEE group had the highest rate among that project's four curricula, and Karnes' GOAL group had a rate between those of the Bereiter-Engelmann and Montessori groups.

Crosstabulations of whether the child had been held back with whether s(he) was in the program were performed. Yates' adjusted chi-squares were calculated for each project.

One should first note from Tables 10 and 12 that the percent of control children held back in grade varies far more among projects than does the percent of control children placed in special education. This is due to two factors: first, the children in the different projects are of different age, and, hence, the projects with older children generally have more children held back because there were more years in which they could have been left back*; and second, in the many school districts where social promotion is practiced, retention in grade is a less frequently used way of dealing with children who are academically failing. (In four of the six projects, the control children were held back less often than they were assigned to special education).

As seen in Table 12, when each project is considered individually only Palmer's project significantly reduced the number of failures. None of the projects which had significantly reduced the rate of assignment to special education also reduced the number of children who were held back. However, the combined results of the seven projects indicate a statistically significant reduction, as do the combined results of the four projects most closely approximating true experiments. Thus, although the overall data suggests that early education can reduce the rate of failure for low-income children, the evidence is not so strong as the evidence for reduction in special education. This could be partially due to the fact that policy with respect to retention varies more widely among school districts than does policy with respect to special education. In any event, the findings must be viewed in light of the fact that four of these projects produced significant or near-significant reductions in the rate of assignment to special education (see Figures 9 and 10).

---

* Assignment to special education tends to occur primarily during the early grades, while retention is more likely to occur at any time.

The Karnes and Miller projects show no significant differences among the different curricula in grade failure. (See Table 13.)

(Note: See Appendix D for explanation of slight differences between these analyses and those reported in "Preliminary Findings of Developmental Continuity Longitudinal Study," May, 1977.)

3. Grade Failure and Special Education in the Micro-Social Learning System Project

The Micro-Social Learning System in Vineland, New Jersey, is discussed separately due to the nature of its control group. Since the program children in this project consisted of the highest risk children in the school district, there did not exist a sufficiently large non-treated group which could be used as a control group. Therefore, it was decided not to compare the high-risk program children to a similar population, but rather to the general school population. A random sample of the previous year's first grade population was chosen. Thus, comparisons could be made to see how the high-risk program children compared to the general school population in Vineland. The general school population included middle class as well as lower class children.

Because the program group had a much higher percentage of Spanish surnamed children than the general school population, an additional thirty-six Spanish children were randomly selected from the prior year's first grade population. These additional children allowed the comparison of Spanish surnamed program children to Spanish surnamed children in the general school population, and, in effect, constituted a control for ethnicity and social class (since most of the Spanish speaking children were from the lower class).

Thus, two comparisons could be made:

(1) program children to the random sample of the general school population (i.e., excluding the additional thirty-six Spanish surnamed children)

(2) Spanish surnamed program children to a random sample of Spanish surnamed children in the general school population (including both the thirty-six additional children and the Spanish children in the first sample).

Employing these two comparisons reveals that the Micro-Social program had an impressive and significant effect on reducing the number of program children who were held back in grade. Even though the program children constituted the high risk population of the school district, they were retrained no more often than the general school population.* As seen in Table 14(a), there was no significant difference between program children and the random sample of the general school population in the number of children held back. The program thus prepared the children to meet minimal school requirements as well as the general school population. This evidence is made even more impressive when one compares the Spanish surnamed program children with the Spanish children in general. As shown in Table 14(b), a significantly lower percentage of the program children were held back (32.8% retained as compared to 62.9%). Thus, the program children were held back at a significantly lower rate than the general school population when even this rough control for ethnicity and social class was employed.

---

* Since the random sample of the general school population had been in school one year longer than the program children, any child in the random sample who was held back only in the last year was counted as never having been held back. This ensured that the comparisons would not be biased. Only one such child was found.

The data on assignment to special education did not reveal as marked an effect, although this may have been due to the fact that the highest risk children were assigned to the program. As seen in Table 15(a), there was not a significant difference between the program children and the general school population; however, the difference was very close to significance -- p=<.066. Further, as seen in Table 15(b), there was no significant difference in assignment to special education between Spanish speaking program children and the general Spanish population.

The fact that the Micro-Social program had a significant effect on retention but not on special education, while other projects had significant effects on assignment to special education but not on being held back, recalls the hypothesis that these programs do affect the percentage of children who meet minimal school requirements and that the question of whether they affect the rate of grade failure or the rate of assignment to special education may be as much a function of the school district's policy on how to deal with underachieving children as it is a function of the specific curricular effects of the early education program.

The Micro-Social findings on the effect of the program on retention, then, are quite impressive since they indicate that, given such an early intervention program, high risk children were able to meet the minimal school requirements just as well as the general school population.

4. Summary of Implications of Special Education and Grade Failure Analysis

The data present very strong evidence that early education for low-income children improves their ability to meet the minimal requirements of the schools they enter.

The combined data from five studies indicate that early education reduces the number of children who were placed in special education classes during their school years. Four of the studies show reductions ranging from 50% to 90% in the number of children assigned to special classes (with al. significance levels below .10). The fifth found no significant differences. The combined results of the five projects showed that the reduction was significant ($p < .01$).

The combined data from seven studies indicates that early education reduces the number of low income children who are held back in grade one or more times ($p < .05$). When these studies are considered separately, however, only Palmer's project showed a significant reduction. None of the projects which significantly reduced the rate of assignment to special education also significantly reduced grade failure.

The Micro-Social Learning System found that high risk children in its program were held back no more often than the general school population. It also found a significant reduction in the number of Spanish speaking children held back.

The combined results from all projects indicate that early education helps low-income children to meet the minimal requirements of their schools. This can consist of reducing the probability of either being assigned to special education classes or being held back in grade. The specific effect of a program could be primarily a function of the predominant policies of local schools for dealing with children who do not meet the minimal requirements. Thus, it appears that early education can result in cost savings be reducing the rate of assignment to special education and/or the rate of grade failure. More importantly, there is now evidence that early education can improve the probability that low income children will be able to perform acceptably in school and not become labeled as failures.

The last section of Part IV will discuss the analysis of school-related responses from the Youth Interview.

Table 10: Effects of Early Education on Special Education for Five Projects

Project	% Program Children in Special Ed.	% Control Children in Special Ed.	Chi-Square	Two-tailed Significance Level	% Reduction in Children in Special Ed.[1]
Closely Approx. Experimental Design					
Gordon[2]	9.4	30.0	3.74	.0524	68.7
Gray[3]	2.8	29.4	5.72	.0168	90.5
Weikart[4]	13.8	27.7	2.77	.0962	50.2

Pooled z score[7] for three projects = 3.46
Pooled significance level[9] = .0006

Project	% Program Children in Special Ed.	% Control Children in Special Ed.	Chi-Square	Two-tailed Significance Level	% Reduction in Children in Special Ed.[1]
Quasi-Experimental					
Levenstein[5]	10.1	39.1	8.17	.004	74.1
Miller[6]	20.4	12.5	0.16	(.6892)	(38.73%)[8]

Pooled z score[7] for five projects = 3.78
Pooled significance level[9] = .0002

Total N = 461

Footnotes:

1 % = [1 - (% Program Children in Special Ed./ % Control Children in Special Ed.)]
2 Majority of children in 3rd grade, Program N=64, Control N=20
3 Majority of children in [illegible] grade, Program N=36, Control N=17
4 Majority of children in 4th grade, Program N=58, Control N=65
5 Majority of children in 3rd grade, Program N=69, Control N=23
6 Majority of children in 7th grade, Program N=93, Control N=16
7 Pooled z = $\frac{\sum z_p}{\sqrt{K}}$ where z_p = z score for significance level of the pth project.
K = number of projects
8 Indicates that a greater percentage of program children are in special education.
9 Two-tailed

Figure 7. Percent of Program and Control Children in Special Education

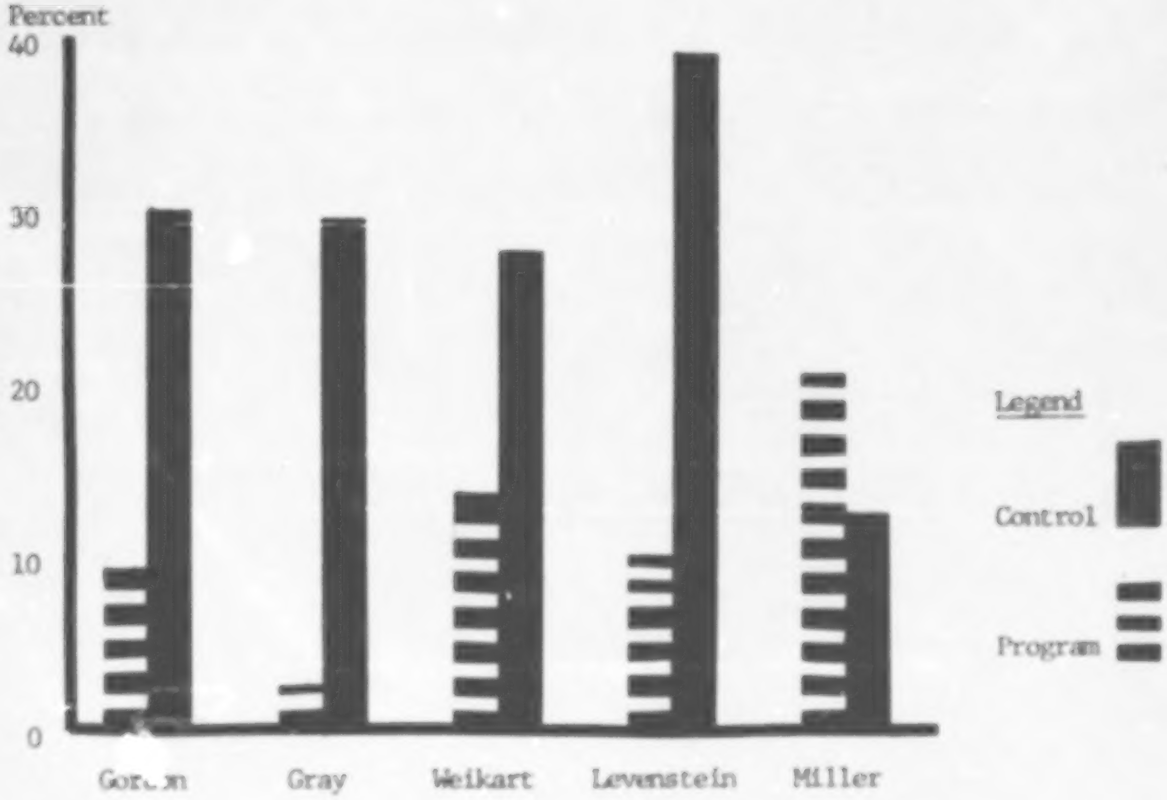


Figure 8. Percentage Reduction in Children in Special Education

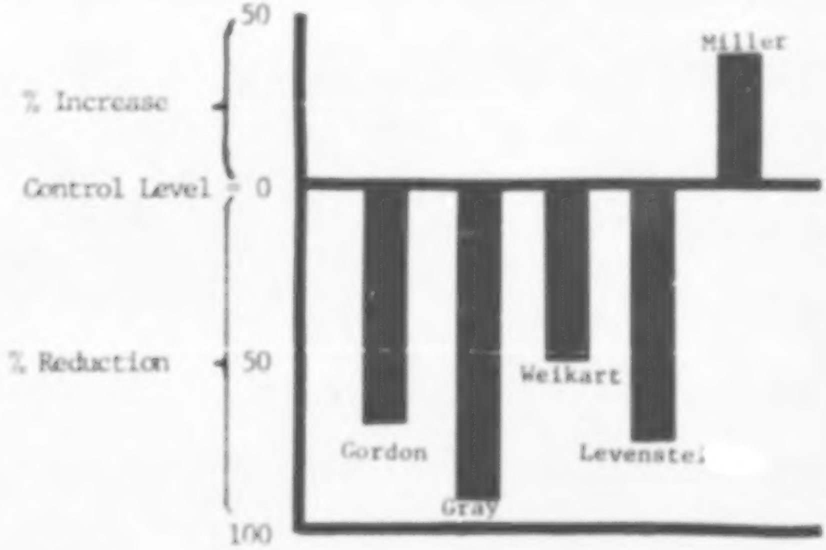


Table 11: The Effect of Different Curricula on Assignment to Special Education*

Louise Miller Project**

	DARCEE	Bereiter-Englemann	Montessori	Traditional	total
Assigned to Special Education	7 (26.9%)	6 (23.1%)	3 (17.6%)	3 (12.5%)	19 (20.4%)
Not Assigned	19 (73.1%)	20 (76.9%)	14 (82.4%)	21 (87.5%)	74 (79.6%)
total	26	26	17	24	93

$X^2 = 1.796$ Not Significant

Merle Karnes Project

	GOAL	Bereiter-Engelmann	Montessori	Traditional	Community Integrated	total
Assigned to Special Education	7 (42.9%)	6 (60%)	4 (30.8%)	2 (25.0%)	6 (54.5%)	25 [illegible]
Not Assigned	12 (57.1%)	4 (40%)	9 (69.2%)	6 (75.0%)	5 (45.5%)	36 [illegible]
total	21	10	13	8	11	63

$X^2 = 3.644$ Not Significant

---

* Comparisons between studies cannot be directly made due to differing availability of special education classes among the various states.

** See appendix B-1.

Table 12: Effects of Early Education on Grade Failure for Five Projects

PROJECT	% Program Children Held Back	% Control Children Held Back	Chi-Square	Two-tailed Significance Level	% Reducation in Children Held Back[1]
Closely Approx. Experimental Design					
Gordon[2]	7.1	9.5	.0116	.9204	25.26
Gray[3]	51.5	66.7	.3196	.5686	22.79
Palmer[4]	22.0	45.0	11.54	.001	51.11
Weikart[5]	3.4	10.8	1.463	.226	68.52
Pooled z score[10] for four projects = 2.64 Pooled significance level[11] = .008					
Quasi-Experimental					
Levenstein[6]	13.2	13.0	.0006	(.9840)	(1.54)[9]
Miller[7]	6.7	0	.3509	(.5552)	——
Zigler[8]	26.6	32.3	.323	.5700	17.65
Pooled z score[10] for seven projects = 1.98 Pooled significance level[11] = .0478					

Total N = 827

Footnotes:

1 % = [1 - (% Program Children Retained/% Control Children Retained)]

2 Majority of children in 3rd grade, Program N = 70, Control N=21

3 Majority of children in 12th grade, Program N = 33, Control N=12

4 Majority of children in 5th grade, Program N = 131, Control N=42

5 Majority of children in 4th grade, Program N = 58, Control N=65

6 Majority of children in 3rd grade, Program N = 68, Control N=23

7 Majority of children in 7th grade, Program N = 105, Control N = 18

8 Majority of children in 7th and 8th grade, Program N = 79, Control N=65

9 Indicates that a greater percentage of program children were retained

10 Pooled $z = \frac{\sum z_p}{\sqrt{K}}$ where zp = z score for significance level of the pth project
K = number of projects

11 two-tailed

Figure 9. Percent of Program and Control Children Held Back a Grade

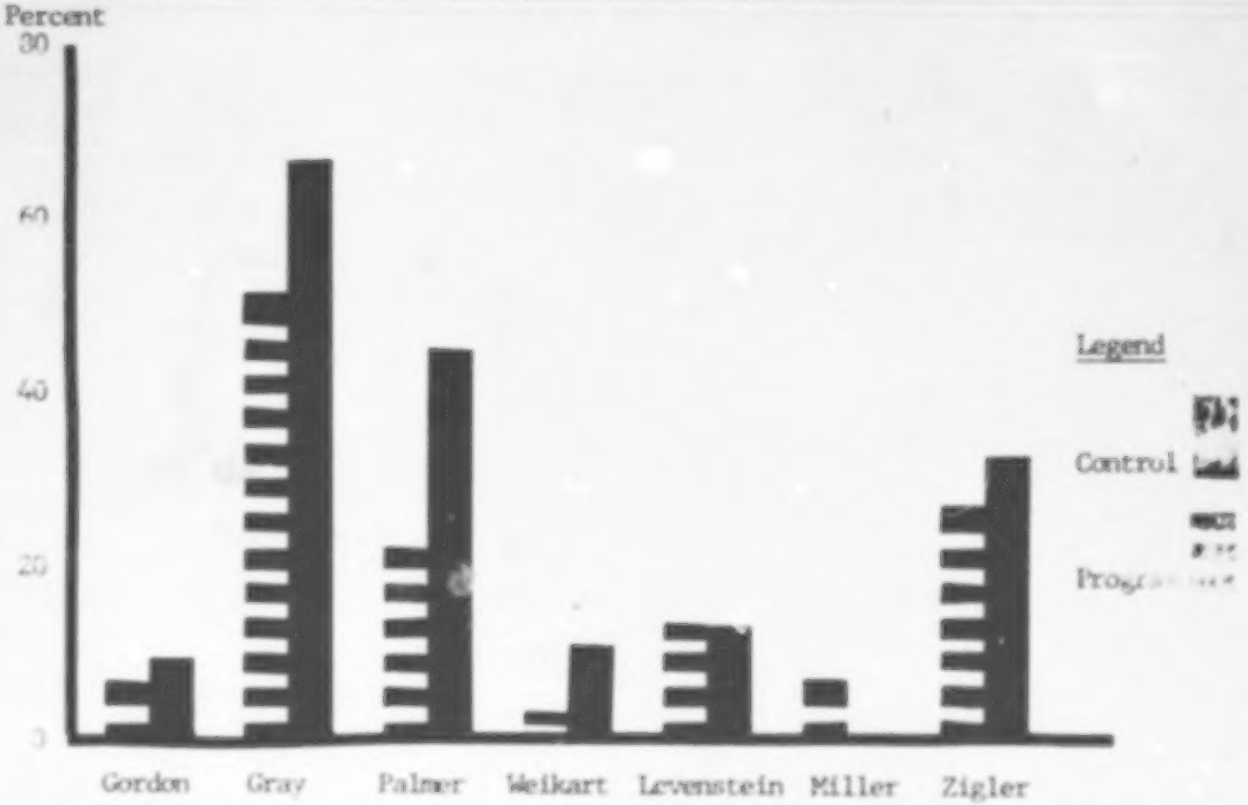


Figure 10. Percentage Reduction in Children Held Back in Grade

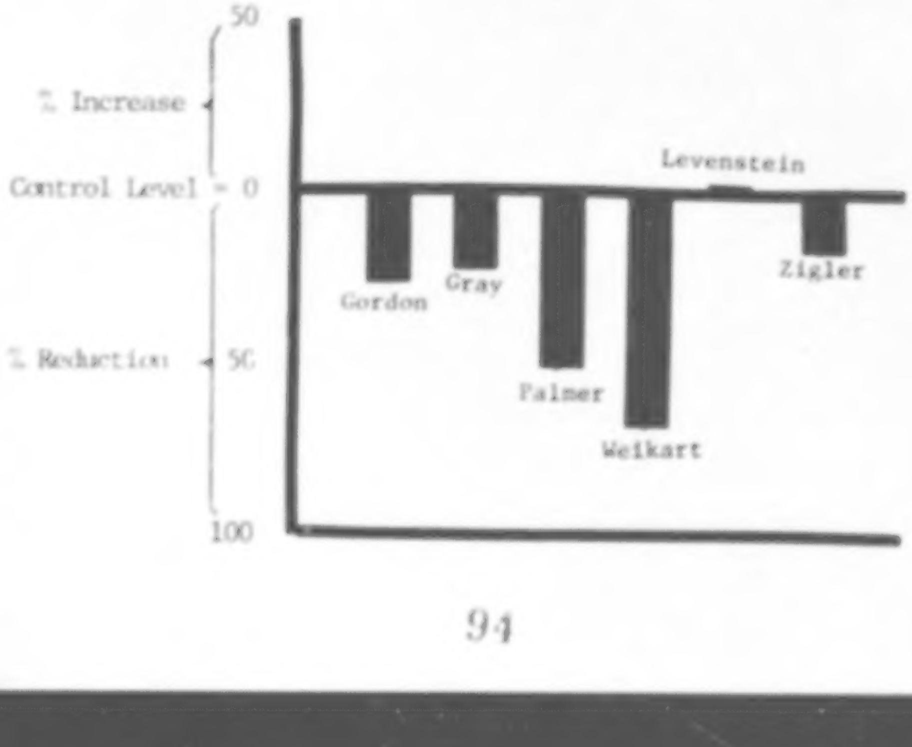


Table 13: The Effect of Different Curricula on Grade Failure

Louise Miller Project

	DARCEE	Bereiter-Engelmann	Montessori	Traditional	total
Held Back	1 (3.7%)	3 (10.0%)	1 (5.0%)	2 (7.1%)	7 (6.7%)
Not Held Back	26 (96.3%)	27 (90.0%)	19 (95.0%)	26 (92.9%)	98 (93.3%)
total	27	30	20	28	105

$X^2 = 1.016$ Not Significant

Merle Karnes Project

	GOAL	Bereiter-Engelmann	Montessori	Traditional	Community Integrated	total
Held Back	4 (18.2%)	2 (20.0%)	3 (21.4%)	0 (0%)	1 (8.3%)	10 (15.4%)
Not Held Back	18 (81.8%)	8 (80%)	11 (78.6%)	8 (100%)	11 (91.7%)	56 (84.6%)
total	22	10	14	8	12	66

$X^2 = 2.632$ Not Significant

Table 14: The Effect of the Micro-Social Learning System on Grade Failure

(a) Program Children vs. General School Population

	Not Retained	Retained	Total
Program N	132	62	194
% of Program	(68.0%)	(32.0%)	(100%)
General Population N	66	35	101
% of General Population	(65.3%)	(34.7%)	(100%)
Total	198	97	295

$\chi^2 = .113$

Significance = 0.74

(b) Spanish Surnamed Program Children vs. Spanish Surnamed School Population

	Not Retained	Retained	Total
Program N	80	39	119
% of Program	(67.2%)	(32.8%)	(100%)
Population N	23	39	62
% of Population	(37.1%)	(62.9%)	(100%)
Total	103	78	181

$\chi^2 = 13.886$

Significance $< .001$

Table 15: The Effects of the Micro-Social Learning System on Assignment to Special Education

(a) Program Children vs. General School Population

	Not in Special Education	In Special Education	Total
Program N	169	27	196
% of Program	(86.2%)	(13.8%)	(100%)
General Population N	95	6	101
% of General Population	(94.1%)	(5.9%)	(100%)
Total	264	33	297

X^2 = 3.39

Significance = .066

(b) Spanish Surnamed Program Children vs. Spanish Surnamed School Population

	Not in Special Education	In Special Education	Total
Program N	104	17	121
% of Program	(86.0%)	(14.0%)	(100%)
Population N	54	8	62
% of Population	(87.1%)	(12.9%)	(100%)
Total	158	25	183

X^2 = 0.0002

Significance = .99

E. Analysis of Schooling Variables from Youth Interviews

The next question is: how far will these children go in school and what factors motivate them? Measures of motivation such as self-concept and educational aspiration are thought to play a role in the number of years of schooling completed. Three variables from the Youth Interview provide information on the educational attainment process. These variables are: Current Status in School (Question 1); Educational Aspiration (Question 2); and Self-Evaluation of Schoolwork (Question 5). A total of 773 Youth Interviews from nine project sites were available on July 1 to investigate differences between program and control groups. The preliminary analyses indicate:

- A slight tendency for more control than program children to drop out of school, and a trend for control children who drop out to leave school at a younger age compared to program dropouts. The differences are not significant, but these findings warrant further investigation when the data collection is completed;

- the children who participated in early education programs do not appear to have educational aspirations that differ from children who did not participate;

- there is, however, evidence in Consortium data that program children rate how they feel they are doing in school higher than children who did not participate in early educational programs. The analysis related to this self-concept finding is discussed below.

1. Self-Evaluation of School Work

Self-concept is a variable often suggested to be related to educational attainment. Enhancement of self-concept, in addition, was often an expressed goal of early intervention programs. Question 5 in the Youth Interview was designed as a measure of self-evaluation related to current self-concept in order to asses an important aspect of the long term impact of early intervention programs.

> Question 5: "How are you doing (did you do) in your schoolwork; that is, overall, not just in one subject? Is your schoolwork ... much better than the others, about the same as others, a little worse than others, much worse than others?"

The response to this question indicated how the sample evaluate their school work compared to their classmates. For this sample (N = 731), 5.6% rated themselves much better than others; 30.5% rated themselves a little better; 54.3% said they were about the same; 8.5% rated themselves worse than others; and 1.1% rated themselves much worse than others.

Chi square analyses show a significant treatment effect for the older subjects. The program group which had participated in Beller, Weikart, Gray and Deutsch's early intervention programs are now at least fifteen years old. These program children are significantly more likely to rate themselves better than others in their school work compared to their controls ($X^2 = 7.812$, $p = < .020$) as shown in Table 16.

Table 16: Self-evaluation of School Work by Program or Control Status (older project sites only)

Self rating	% Control	% Program
Better than others	33.6	48.5
About the same	56.6	47.7
Worse than others	9.8	3.8
N of cases	122	132

$X^2 = 7.812$, $p = < .020$

Looking at the four project sites individually, they are all in the predicted direction but only Deutsch is significant ($X^2 = 5.94$, $p =< .05$). The program vs. control comparison is not significant for the younger subjects and there are no significant sex differences.

Taken with the other data on special education placement and failure to be promoted, this self-rating data would appear to be a relatively accurate assessment. There is no reason to expect that the program group would be more likely than the controls to say what they think the interviewer would like to hear, and it is assumed that the interviewers for the most part were blind to the program or control status of the respondent. This program effect appears in the sample when the subjects reach high school -- a time when decisions about educational goals are critical.

In sum, further analyses of Youth as well as Parent Interview responses can serve as an important data source for the investigation of attitudinal and interpersonal variables that mediate the effect of early education and later educational outcomes.

Because of the complexity of analysis of the interview data, and because it does not contribute particularly to the main effects reported in this paper, we have decided to delay its full treatment until the whole body of data has been received.

Several studies experienced unavoidable delays in data gathering while waiting for school districts to decide upon the adequacy of their informed consent and privacy procedures. All finally did get access to school data, but in one study (Beller) data collection will not be completed until later in the fall, and in several other cases, data could not be collected in time for this required report.

PART V

ANALYSIS OF DELIVERY SYSTEMS

A. Parental Evaluations of Programs

As part of the Parent Interviews, parents of the experimental and of some of the control children were asked to evaluate the programs their children had been in ten years earlier. The responses of these 684 parents were quite positive, and these final results confirm the preliminary findings reported in May on several hundred fewer respondents.

While in some studies control parents were asked to evaluate "programs" in order to maintain the blind condition of the interviewer, the primary results to be reported here are from interviews with parents of experimental children.

In response to the question, "Was the program a good thing for your child?" many more parents answered "yes" than "no" or "don't know." As may be seen in Table 17, all of the parents whose children had been in home-based programs answered the question positively, as did 93.4% of the parents of children from center-based programs, and 87.8% of the parents of children who had been in combination programs. There was more uncertainty, however, in the center and combination delivery systems where nineteen parents (4.3%) and seventeen parents (10.9%) respectively answered that they did not know whether the program had been a good thing or not. Ten (2.3%) of the center-based program parents and two (1.3%) of the combination program parents stated that the programs had not been good for their children. While these last negative findings might be of concern to program administrators, they are extremely small percentages, and may indeed be indicators of the validity of the interview in that this question evidently was not (at least to some parents) one that intrinsically demanded a positive parental response.

Table 17: Responses to Parental Interview Question "Was the Program A Good Thing for Your Child?" By Delivery System

N=684

	Center		Home		Combination	
Answer	N	%	N	%	N	%
Yes	412	93.4	87	100	137	87.8
No	10	2.3			2	1.3
Don't Know	19	4.3			17	10.9
Total	441	100	89	100	156	100

$x^2 = 18.055$ $p = <.0012$

The chi-square performed for relationships between the variables of answers to the question and delivery system reveals a statistically significant relationship, with home programs being most likely to have parents who answered the question positively.

In order to determine the basis of their judgments about the program's value, the parents were asked what they liked best. While there was considerable variance across delivery systems (which will be discussed below), overall the best liked category related to cognitive aspects of the program. This group of responses included mentions of learning with toys, etc. Next most popular overall were program characteristics (22.3%), including teacher-child ratio, length of program, materials, etc. Also approved of at similar levels were the total program (13.6%), parental aspects (13.1%), and social aspects (15.4%). Additional services (such as medical care) accounted for 3% and "nothing" (usually meaning "nothing in particular") accounted for 1.2%.

Table 18: Answers to the Question "What Did You Like Best About the Program Your Child was In?" By Delivery System

N=605

	Center		Home		Combination	
	N	%	N	%	N	%
Cognitive Aspects	134	34.3	17	20.5	40	30.5
Program Characteristics	85	21.7	29	34.9	22	16.0
Social Characteristics	80	20.5	2	2.4	11	8.4
Total Program	43	11.0	15	18.1	24	18.3
Parental Aspects	29	7.4	19	22.9	31	23.7
Additional Services	15	3.8	1	1.2	2	1.5
Nothing	5	1.3			2	1.5
Total	391		83		131	

Of particular interest are the differences among the delivery systems on social, parental, and program characteristics. As would be hypothesized, parents of children who had been in center-based progams (which had the salient feature of groups of children interacting with each other and with adults) were much more likely to identify social aspects of the program as best liked than were the parents of children who had been in the home or combination programs. Conversely, parents of children who had participated in home-based or combination programs were nearly three times as likely to identify parental aspects as best liked as were parents of children who had been in center-based programs.

Even so, in no delivery system did the parents choose parental benefits more often than they did direct benefits to their children. The most frequently mentioned category for the home-based programs was program characteristics and for the combination and center it was cognitive benefits.

The last category "nothing" should not be interpreted as meaning that the parents did not like anything about the program. Rather, analysis of the answers for the seven parents who answered this way reveals that the answer apparently meant that they liked everything about the program and not one thing in particular. Only one of the seven said that she did not like the program; two others said that they couldn't remember anything about the program except that it was nice.

To obtain more critical evaluations of the programs, the parents were asked what they did not like and how they would have changed the program to make it better. By a wide margin, most parents responded that there was not anything they did not like. By delivery system, 84.1%, 83.1%, and 89.6% of the center, home, and combination program parents responded in this manner. In other categories 8.7% overall answered that they did not like some of the characteristics of the program (like the teaching philosophy).

As to what they would change about the programs, 57.3% overall said that they would change nothing, 13.8% responded that they would change program characteristics such as the length or frequency of the program. Other categories were quite low in responses.

The parents were asked if they liked the location of the program. (A preceding statement explained that sometimes programs for young children are conducted in centers and sometimes a home visitor brings activities to the home.) Parents in both home and center programs preferred the location that they had utilized: 98.8 % and 96.6% of the center and home program parents respectively said they liked the location of the program. Only one parent from a center-based program and three from home-based programs did not like the location of the program, as may be seen in Table 19.

Table 19: Responses to the Question "Did You Like Having the Program in a Home/Center?" By Delivery System

	Home		Center	
	N	%	N	%
Yes	84	96.6	422	98.8
No	3	3.4	1	.2
Don't Know	0	0	4	.9
Total	87	100	427	99.9

For the combination programs, the parents were asked if they liked having the program in both locations; 88.7% (118 parents) responded that they did like having both home and center locations and 11.3% (fifteen parents) said that they did not like them. When asked if they preferred either the home or center component of the program, 58% (eighty parents) had no preference, 31.2% (forty-three parents) preferred the center program, and 10.9% (fifteen parents) preferred the home program.

When parents were asked why they liked the location of the program, distinct differences emerged between the home and center-based studies. In contrast to what they liked best, the parents of children in home-based studies most frequently said that they liked parental aspects (76.2%). For the center-based studies, the most frequent response was that they liked everything about the program -- or that they "just liked it" (27.2%). Another 25.9% of these parents answered that they liked the social aspects of the center location.

Similarly, when the combination program parents were asked what they liked best about the double locations of those programs, 56.8% responded with mention of the parental aspects, while 18.9% answered with the general "everything" or "I just liked it," "It was fine."

In summary, the parents stated that they considered the programs to have been valuable to their children in a variety of developmental ways. There was little that they did not like about the programs, and little that they would have changed. Generally, they liked the location of the programs and the aspects of it that made the location unique; that is, parents of home-based and combination program children liked parental aspects of the programs and parents of center-based program children liked the total program best.

Generally it may be deduced that these programs were highly successful in obtaining parental satisfaction with their efforts.

B. Program Variables Analysis

Are some intervention programs significantly more effective than others? If so, what kinds of programs are most effective? We have just begun to address this important set of questions.

It seems reasonable to approach this topic in four steps:

1. Determine whether there are reliable differences among the programs, as measured by the difference between experimental group and control group means on IQ's or other dependent variables.

2. Determine whether these apparent differences among programs (if any such differences are found) might be due to the fact that some projects had methodological biases favoring experimental groups, and other projects had methodological biases favoring control groups. If such biases existed, then statistically reliable differences might appear in the apparent effectiveness of the various programs, even if there were no real differences in program effectiveness.

3. If differences among project results appear to be due to real differences in program effectiveness, the next step would be to determine whether these differences are caused by identifiable features of the programs, such as delivery system, staff training, etc., or whether they might be due merely to idiosyncratic features of individual programs that cannot be replicated in the future, such as personal effectiveness of individual teachers or supervisors. This would involve demonstrating some consistency across projects, in the types of projects which yield superior results.

4. If the results of step 3 are positive, the next step would be to see whether program variables are confounded with each other. If

superior programs consistently have feature A, do they also have feature B so that it is not clear which feature produced the superiority?

Of the four-step program which would be necessary before any useful conclusions can be drawn from this type of analysis, only the first step has even been approached seriously, and even there we have used only one dependent variable of perhaps minor interest—Stanford-Binet IQ measured immediately upon completion of the program. Only one statistically reliable difference has been detected thus far--programs having medium or high degree of teaching structure appeared to have greater effects on immediate posttest IQ than programs with low teaching structure. Even for this finding we have not yet completed step 2 above--determining whether this result might be caused by biases favoring the experimental groups in some projects. Preliminary inspection indicates that even if this finding survives a stage-2 analysis, it will probably not survive a stage-3 analysis. Thus the findings in this area so far--which involve only immediate posttest IQ--are essentially negative.

This does not mean that program characteristics have no effect on development. It may mean that there were too many ambiguities in our data to draw definite conclusions. Or it may be that clear differences will emerge for other dependent variables. Or it may mean that the projects we studied were all managed too skillfully for differences to appear, even if differences might appear for less carefully managed projects. To take only one example, if it makes no detectable difference whether children were taught in groups of two or five children, there may nevertheless be important differences between groups of two or five on the one hand, and groups of twenty or twenty-five children on the other.

The stage-1 analysis proceeded as follows. First, a measure of the effectiveness of each intervention program was computed. At first we considered using the simple difference between the means of the experimental and control groups for each project. However, because some subject populations were more homogeneous than others, and because different projects used different IQ posttests, it seemed desirable to adjust these for the standard deviation of each project's IQ scores. To do this, we computed a point-biserial correlation for each project, between immediate posttest IQ and treatment-control status. This correlation is essentially a measure of the effectiveness of the project, as measured by immediate posttest IQ scores, standardized for the standard deviation of IQ's in the project. These effectiveness measures are shown in Table 20.

A list of variables relevant to all programs, such as adult-child ratio, degree of structure in the teaching method, etc., was compiled using many of the variables identified by Gordon, et al.* and by Goodson and Hess.** Based on written reports and raw data, each variable was assessed and assigned a value for each program (see Table 21). For example, degree of structure in the teaching method was determined to be high, moderate, or low for each study. This listing was then sent to each investigator for review and corrections before it was used in the analysis. Not all studies were included in the analysis because lists were not returned in time for the preparation of this report. In addition, two studies with the greatest diversity in program (Karnes' Curriculum Comparison and Weikart's Curriculum Demonstration) could not be included in this analysis because they did not have control groups.

---

* Gordon, et al. "Research Report of Parent Oriented Home-based Early Childhood Education Program." Institute for Development of Human Resources, May, 1975.

** Goodson, Barbara Dillon and Robert Hess. "The Effects of Parent Training Programs on Child Performance and Parent Behavior." Stanford University. undated mimeograph.

For the analysis, each variable was forced into dichotomous categories. For example, Target Group was dichotomized as children only versus parents and children. Each study (or its relevant groups) was then assigned to the appropriate category. In those cases in which a study was divided between groups, separate effectiveness correlations were calculated for each part of the study.

Using the following formula, a z score was calculated for each program variable. This formula was suggested by Richard Darlington and incorporates the fact that the standard error of a low product-moment correlation is about 1/n.

$$z = \frac{\sum c_i v_i}{\sqrt{\sum \frac{v_i^2}{n_i}}}$$

c_i = the effectiveness correlation

v_i = a weight (the sum of all of which must equal zero)

n_i = the number of subjects in a study, i

A total of twenty-five comparisons were made, involving fifteen different program variables. The uncorrected p-values for these comparisons are shown in Table 22. Each of these p-values should be multiplied by about twenty-five (the actual figure may vary slightly) to obtain a significance level, to correct for the fact that they were selected post hoc from among the twenty comparisons. When this step is taken, only one of the comparisons is still significant: for low teaching structure vs. medium or high teaching structure we have a significance level of 25 x .0014 - .035. It must be remembered that the effective sample size for this result is the total number of children in all projects together, or 2060. A result which is significant at only the .035 level in such a large sample should not be taken too seriously.

In summary, the analysis outlined at the beginning of this section involves four "screens" which a finding must pass before it can be taken seriously. So far, only one finding has passed even the first of these four screens, and its outlook for passing the other three screens seems dim. However, we have studied only immediate posttest IQ, and more positive results might be found if other variables are studied.

Table 20: Correlations Between Immediate Posttest IQ and Treatment-Control Status

	Sample Size	Correlation
Beller	163	.2073
Deutsch	740	.3564
Gordon	186	.1682
Gray-Early Training Project	87	.4982
Gray-Family-Oriented Home Visitor	51	.2454
Levenstein	150	.3082
Miller	231	.0091
Palmer	197	.1289
Weikart Perry Preschool Project	120	.2179
Weikart Carnegie Infant Study	36	.0890
Zigler	99	.0410

Table 21: Matrix of Program Descriptions

	Beller	Deutsch	Gordon
Age at beginning of intervention**	*Group 1 = 4 yr. Group 2 = 3 yr.	4 yrs.	Groups 1,2,4,5 = 3 mo. groups 3,6 = 1 yr. Group 7 = 2 yrs.
Length of Intervention in Years	Group 1 = 2 yrs. Group 2 = 1 yr.	5 yrs.	Group 1 = 2 yrs. 9 mo. Groups 2,3,4 = 1 yr. 9 mo. Groups 6,7 = 1 yr.
Length of intervention in mo/yr.	9 mos.	9 mos.	12 mos.
Goals stated for parents and/or children **	Children	Children	Parents and children
Geographic location by HEW regions**	Region 3 (Mid-Atlantic)	Region 2 (Northeastern)	Region 4 (Southeastern)
Child Condition at Entry**	Normal	Normal	Normal
Delivery system** Site	Center	Center	Home; Home and Center
Adult-Child ratio**	Nursery School: 1-7.5 Kindergarten: 1-30	Nursery School: 1-7.5 Kindergarten and above: 1-12.5	Home: 1-1 Center
Staff Qualifications	Nursery professional teacher Paraprofessional aide Kindergarten-Professional teacher	Professional teacher Paraprofessional aide	Paraprofessional
Staff training	In-Service	Pre-Service and In-Service	Pre-Service and In-Service
Progam goals	Cognitive Affective Language	Cognitive, Language, Perceptual, Self-Image	Maternal, Competence Child-Psychomotor, Cognitive, Affective
Degree of structure in teaching activities	Moderate	Moderate	Moderate
Specificity in instruction by parents or teachers	Moderate	Moderate	Moderate
Emphasis on language training	Moderate	High	Moderate
Philosophical orientation	Traditional socio/emotional	Cognitive and socio/emotional	Cognitive
Parental Involvement	Moderate	Moderate	High
Number of hours of intervention per year	Nursery and Kindergarten: 720 hrs/yr	Nursery and Kindergarten: 720 hr/yr School grades: 1350 hr/yr	Home: 52 hr/yr Center: 208 hr./yr
Hours per year x adult-child ratio	Nursery and kindergarten: 95.76	Nursery and kindergarten: 93.6 School grades: 108	Home: 52 Center: 104
Child-child ratio	Nursery: 1-15 Kindergarten: 1-30	Nursery: 1-15 Kindergarten and school gardes: 1-25	Home: 0 Center: 1-5

---

* See last section of Table 20 for key to Group Codes by project site.

** These variables are taken from Gordon or Goodson and Hess.

cont.	Gray Early Training Project	Gray - Family Home Visitor
Age at beginning of intervention**	*Group 1 = 3.8yrs	16-22 mo.
Length of intervention in Years	Group 1 = 2 yrs. Group 2 = 1 yr.	Groups 1 & 3 = 2 yrs. Groups 2 & 4 = 1 yr.
Length of intervention in mo/yr.	12 mos.(10 week summer school 9 mos. home visits)	
Goals stated for parents and/or children**	Both	Both
Geographic location by HEW regions**	Region 4 (Southeastern)	Region 4 (Southeastern)
Child Condition at Entry**	Normal	Normal
Delivery system** Site	Home and Center	Home
Adult-Child ratio**	1-4 summer school 1-1 home visits	1-2
Staff Qualifications	Professional teacher, paraprofessional aides & home visitors	Professional and paraprofessional
Staff training	Pre-service and In-service	Pre-service and In-service apprenticeship
Program goals	Cognitive attitudinal	Cognitive affective, parental competence
Degree of structure in teaching Activities	High	Moderate
Specificity in instruction by parents or teachers	High	Moderate
Emphasis on language training	Moderate	Moderate
Philosophical orientation	Cognitive	Cognitive and socio-emotional
Parental Involvement	High	Moderate
Number of hours of intervention per year	240 hr./yr.	24 hr./yr
Hours per year x adult-child ratio	100	24
Child-child ratio	5 for summer school 0 for home visits	0

---

* See last page of Table 19 for key to Group Codes by project site.

** These variables are taken from Gordon or Goodson and Hess.

Table 21 cont.	Karnes	Miller
Age at beginning of intervention**	4 yrs.	4 yrs.
Length of intervention in Years	1 yr.	4 yrs. - intervention & regular school; 5 yrs. - intervention & work-spend
Length of intervention in mo/yr.	8 mo/yr	9 mos/yr
Goals stated for parents and/or children**	Children	Children
Geographic location by HEW regions**	Region 5 (Midwestern)	Region 4 (Southeastern)
Child Condition at Entry**	Normal	Normal
Delivery system** Site	Center	Groups 2,3,4 = center 1/2 of group 1 = center, 1/2 of group 1 = combination
Adult-child ratio**	*Groups 1,2,3 = 1-5 group 6 = 1-6 group 7 = 1-10	Groups 1,2,3 = 1-8 group 4 = 1-7
Staff Qualifications	Professional & paraprofessional	Professional and paraprofessional
Staff training	In-service	Pre-service and In-service
Program goals	Group 1 - language cognitive, affective group 2,3 - cognitive group 6 - cognitive, self-development group 7 - language, cognitive, affective	Group 1,2 - cognitive, language, achievement motivation, group 3 - cognitive, sensory, daily living, character group 4 - congitive, emotional, social, physical
Degree of structure in teaching activities	Groups 1,7 - minimal groups 2,3,6 - high	Groups 1,2,3 - high group 4 - moderate
Specificity in instruction by parents or teachers	Groups 1,7 - moderate groups 2,3,6 - high	Groups 1,2,3 - high group 4 - low
Emphasis on language training	Groups 1,2,3,7 - high group 6 - low	Groups 1,2 - high group 3 - low group 4 - moderate
Philosophical orientation	Group 1,7 - traditional groups 2,3,6 - behaviorial	Groups 1,2 - cognitive, behaviorial group 3 - cognitive, personal; group 4 - traditional socio/emotional
Parental Involvement	Moderate	Groups 2,3,4,& 1/2 of 1 - minimal; 1/2 of group 1 - high
Number of hours of intervention per year	[illegible] hr/yr	Group 1 - 1206 hr/yr groups 2,3,4 - 1160 hr/yr
Hours per year x adult-child ratio	72	Groups 2,3& 1/2 of group 1 - 146.25; group 4 - 166.14; 1/2 of group 1 - 182.25
Child-child ratio	5	Groups 1,2 - 8 group 4 - 7 group 3 - 2

* See last page of Table 19 for key to Group Codes by project site.

** These variables are taken from Gordon or Goodson and Hess.

cont.

	Levenstein	Palmer
Age at beginning of Intervention**	2 yrs.	Groups 1,2,6,7 = 2 yrs. Groups 4,5 = 3 yrs.
Length of Intervention in years	*Groups 1,16 = 9 mo. Groups 5,7,8,12,14 15 = 1 yr., 9mos.	Groups 1,2,4,5 = 1 yr. Groups 6,7 = 2 yrs.
Length of Intervention in mo/yr.	7 mos/yr	8 mos.
Goals stated for Parents and/or Children**	Parents & Children	Children
Geographic location by HEW regions **	Region 2 (Northeastern)	Region 2 (Northeastern)
Child condition at entry**	Normal	Normal
Delivery system** Site	Home	Center
Adult-Child Ratio**	1-1	1-1
Staff Qualifications	Professional, Volunteer, Paraprofessional	Professional and paraprofessional
Staff training	Levenstein Preservice and Inservice	Palmer In-service
Program goals	Cognitive, affective, language	Cognitive
Degree of structure in teaching activities	High	Groups 1,4,6 - High Groups 2,5,7 - low
Specificity in instruction by parents or teachers	High	Groups 1,4,6 - High Groups 2,5,7 - low
Emphasis on language training	High	Moderate
Philosophical orientation	Cognitive and Traditional socio/emotional	Cognitive
Parental involvement	High	Moderate
Number of hours of intervention per year	46 hr/yr (long year) 7 hr/yr (short year)	45 hrs.
Hours per year x adult-child ratio	46 long year 7 short year	45
Child-child ratio	0	1/0

* See last page of Table 19 for key to Group Codes by project site.

** These variables are taken from Gordon or Goodson and Hess.

Table 21
cont.

	Weikart Perry Preschool Project	Weikart Curriculum Demonstration	Woolman
Age at Beginning of intervention **	*Group 1 = 3 yrs. Group 2 = 4 yrs.	3 yrs.	4 yrs.
Length of Intervention in Years	Group 1 - 17 mos. Group 2 = 8.5 mos.	2 yrs.	2 yrs.
Length of Intervention in mo/yr.	8.5 mos/yr.	8.5 mos.	12 mos.
Goals stated for parents and/or children*	Children	Children	Children
Geographic location by HEW regions**	Region 5 (Midwestern)	Region 5 (Midwestern)	Region 2 (Northeastern)
Child Condition at Entry**	Functionally Mentally Retarded	Functionally Mentally Retarded	Educably Mentally Retarded
Delivery System** Site		Combination	Center
Adult-Child Ratio**	Home: 1-1 Center: 1-6	Home: 1-1 Center: 1-5	1-15
Staff Qualifications	Professional, para-professional and volunteer	Professional and paraprofessional	Professional and Paraprofessional
Staff Training	Weikart Perry Preschool In-Service	Wekart Curriculum Demonstration In-Service	Woolman in-service
Program Goals	Cognitive, language affective	Cognitive	Language, Cognitive, Social, Achievement Motivation.
Degree of structure in teaching activities	Moderate	Groups 1 & 2 - high Group 3 - low	Moderate (1/2 high, 1/2 low)
Specificity in instruction by parents or teachers	Moderate	Groups 1 & 2 - high Group 3 - low	Moderate
Emphasis on language training	Moderate	Groups 1 & 3 - high Group 3 - moderate	High
Philosophical Orientation	Cognitive	Group 1 - Cognitive Group 2 - Behavioral Group 3 - Traditional	Cognitive, Social
Parental Involvement	High	High	Moderate
Number of hours of Intervention per year	504 hrs.	450 hrs.	600 hrs./yr.
Hours per year x adult-child ratio	128.7	198	39.6
Child-child ratio	1/5	1/5 in center 1/1 in home	1/15

* See last page of Table 19 for key to Group Codes by project site.
** These variables are taken from Gordon or Goodson and Hess.

Table 21
cont.

	Weikart Carnegie Infant Study	Zigler
Age at Beginning of Intervention**	3, 7, or 11 mos.	4 yrs
Length of Intervention in Years	16 mos.	5 yrs. for follow-through group
Length of Intervention in mo/yr.	12 mos. 1st year 4 mos. 2nd year	9 mos.
Goals stated for parents and/or children**	Both parent and children	Children
Geographic location by HEW regions**	Region 5 (Midwestern)	Region 1
Child Condition at Entry**	Normal	Normal
Delivery system** Site	Home	Center
Adult-Child ratio**	1-1	1-5 for Head Start 1-10 for follow Through
Staff Qualifications	Professional	Professional and Para-professional
Staff Training	Weikart Carnegie Infant In-Service	Zigler In-Service
Program Goals	Cognitive, Mother as Teacher	Cognitive, Social, Emotional, Physical
Degree of structure in teaching activities	Moderate	Low - Head Start Moderate - Follow Through
Specificity in instruct instruction by parents or teachers	Moderate	Moderate - Follow Through Low - Head Start
Emphasis on language training	Moderate	Moderate
Philosophical Orientation	Cognitive	Traditional, Socio-Emotional
Parental Involvement	High	Moderate
Number of hours of intervention per year	48-72 hrs/yr.	1160 hrs./yr.
Hours per year x adult-child ratio	72	108 - Head Start 126 - Follow Through
Child-child ratio	1/0	1/5 - Head Start 1/10 - Follow Through

** These variables are taken from Gordon or Goodson and Hess.

Table 21
cont.

Group Codes for Program Variable Descriptions

Beller

Group 1 = Entered nursery school at 4
2 = Entered kindergarten at 5
3 = Entered first grade at 6

Deutsch

All groups the same

Gordon

Group 1 = Experimental first, second and third years
2 = Experimental first and second years, control third year
3 = Control first year, experimental second and third years
4 = Experimental first and third years, control second year
5 = Experimental first year, control second and third years
6 = Control first and third years, experimental second year
7 = Control first and second years, experimental third year
8 = Control all three years

Gray: Early Training Project

Group 1 = Experimental at age 3.8 years
2 = Experimental at age 4.8 years
3 = Random local control
4 = Distal control

Gray: Family-Oriented Home Visitor

Group 0 = Control
1 = Extensive home visiting for 2 years
2 = Extensive home visiting for 1 year
3 = Materials only for 2 years
4 = Materials only for 1 year

Karnes

Group 1 = Traditional Curriculum
2 = Bereiter-Engelmann
3 = GOAL program
6 = Montessori
7 = Community Integrated

Levenstein

Group 1 = 1 year treatment
5 = 2 years treatment
7 = 1 and 2 years D
8 = 1 and 2 years D Materials only
12 = 2 years materials only
14 = 1 full year, 1 short year
15 = 1 full year, 1 short year, materials only
16 = 1 year control (visitor), 1 year treatment
17 = Control, 1967 (2 & 3 yrs. old)
18 = Control, 1967 (4 yrs. old)
19 = Control, after only

Table 21
cont.

Miller

Group 1 = DARCEE
2 = Bereiter-Engelmann
3 = Montessori
4 = Traditional
5 = Original Control

Palmer

Group 1 = Training at 2 years for one year, concept training
2 = Training at 2 years for one year, discovery training
3 = Control
4 = Trained at 3 years for one year, concept training
5 = Trained at 3 years for one year, discovery training
6 = Training at ages 2 and 3, concept training
7 = Training at ages 2 and 3, discovery training

Weikart Perry Preschool Project

Group 0 = Control
1 = Experimental from ages 3-5
2 = Experimental from ages 4-5

Weikart Curriculum Demonstration Study

Group 1 = Cognitively-oriented curriculum
2 = Bereiter-Engelmann
3 = Traditional

Weikart Carnegie Infant Study

Group 1 = Experimental
2 = Control
4 = Post-tested only controls

Woolman

All groups the same

Zigler

Group 0 = No follow-through
1 = Follow-through

Table 22: Z-Score Computation for Between-Program Variable Categories

Variable	Contrast	z	p
Geographic Region	Northeast vs. Southeast & Midwest	- .534	.5962
	Northeast and Midwest vs. Southeast	- .5652	.5754
Child Condition at Entry	Normal vs. Retarded	.1323	.8966
Target Group (Goals Stated for Whom)	Children Only vs. Parents and Children	.1646	.8728
Length of Intervention in Months/Years	Over 9 Months vs. Less Than 9 Months	.7478	.4592
Delivery System	Home vs. Center and Combination	- .0765	.9442
	Combination vs. Home and Center	1.5637	.1188
	Center vs. Home and Combination	-2.0850	.0376*
Staff Qualifications	Paraprofessional and Volunteer only vs. Paraprofessional and Professional/and Professional only	.6034	.5486
	Professional only vs. Paraprofessional, Professional and Volunteer	1.0340	.3030
Child Group Size	One vs. 5 or more children	- .5785	.5686
Emphasis on Language	Medium vs. High and Low	.9763	.3320
	High vs. Medium and Low	1.2690	.2076
	Low vs. Medium and High	-1.4352	.1528
Philosophical Orientation of Program	Cognitive and Behavioral vs. Traditional	.0625	.5352
Adult-Child Ratio	1:1 vs. 1:1 and 1:>1	.3101	.7566
	1:>1 vs. 1:1 and 1:>1	-1.4962	.1362
Teaching Structure	High vs. Medium and Low	1.7477	.0818
	Medium vs. High and Low	.3031	.7642
	Medium and High vs. Low	-3.2088	.0014*

* uncorrected p values $<.05$

Variable	Contrast	z	P
Parental Involvement	Medium and High vs. Low	2.3547	.0118*
	High vs. Medium and Low	-1.2329	.2186
Staff Training	Pre-service and In-service vs. In-service Only	-2.0820	.0376*
Intensity of Intervention	Less than 100 hours per year vs. More than 100 hours per year	-1.6199	.1074
Adult Intensity Ratio	Less than 50 Potential Adult Hours of Contact vs. More than 50 Hours	.4196	.6744

* uncorrected p-values < .05

PART VI

CONCLUSIONS AND IMPLICATIONS

A. Conclusions

If one accepts all of the assumptions inherent in the analyses, then the following conclusions seem reasonable:

1) Infant and preschool services improve the ability of low income children to meet the minimal requirements of the schools they enter. This effect can be manifested in either a reduced probability of being assigned to special education classes or a reduced probability of being held back in grade. Either reduction constitutes a substantial cost reduction for the school system.
2) Low income adolescents who received early education rate their competence in school higher than comparable adolescents who did not have preschool education.
3) As measured by the Stanford-Binet and the WISC tests preschool programs produce a significant increase in the intellectual functioning of low-income children at least during the critical years of the primary grades in school.

Probably the most important finding is that low income children who received early education are better able to meet the minimal requirements of their school. Results on IQ tests indicate that this may be due to an increase in intellectual skills which lasts through the very important primary grades in school. Additionally one might hypothesize that the reduction in rate of assignment to special education or of being held back in grade is due in part to increased parental concern and competence to deal with the school system - an indirect effect of parental sensitization by the preschool experience of their children. This hypothesis is certainly supported by the fact that all of the projects which found significant effects on special education either were home visiting programs or had some visiting components. The two projects which found significant effects on grade failure had somewhat lesser degrees of parental involvement. At this point, then, the evidence does not differentiate between the relative contribution to the effects on school performance of increased intellectual skills in the child and increased parental concern with the child's learning. The positive effect of early education on school performance is now fairly well established, but the mechanism by which it has this effect is still open to conjecture.

The evidence also suggests that, with respect to school performance, there is as of now no indication of a "magic age" at which early intervention is most effective. Projects having significant effects on assignment to special education were Gordon (education from birth to two years of age), Gray ETP (four and five years of age) and Levenstein (two and three years of age). Projects having significant effects on grade failure were Palmer (two and three years of age) and Woolman (four years of age). Further, programs which had significant effects ranged in duration from eight months (Palmer) to up to three years (Gordon). The most that can be said is that, at this point, there is no indication of a most effective age or length of intervention. However, given the relatively small number of programs investigated and the relatively large number of confounding factors (geographic region, cohort, school policy, etc.), it is possible that such magic points exist, but are not discoverable without a more rigorous, large scale experiment designed specifically to answer those questions.

In sum, the most important conclusion which has been reached to date from this body of data is that well-run early education programs can in some way improve the ability of low income children to meet the requirements of their schools.

B. Implications and Recommendations for National Policies and Programs

In examining both the statistically significant findings, and the trends which appear across these studies, certain policy implications seemed clear to us. While further analyses and other data will be necessary to fully document these implications, we felt it important that they be listed now, even in their somewhat tentative state.

1. Day Care

These findings indicate that the failure to require a deliberate, well planned curriculum for young children in federally supported day care programs is likely to cost more money in later special education expenditures than would be saved in day care costs. We recommend that the educational requirement not only be restored to the Federal Requirements, but should be made explicit.

2. Program Delegation

While it is difficult to generalize across all public schools, in none of the public schools who sponsored preschool programs included in this study were parents involved as teachers of their own preschool children. Neither did they mount home-based services, nor accept children under four years of age. Further -- there are exceptions -- most public schools have rigidly defined and highly structured curricula, and the typical supervisory structure of a public school does not permit very much leeway in structure. The tendency to adopt uniform policies and philosophies in districts mitigates against the capacity of many school districts to be responsive to individual and subcultural needs in the way a fifteen-child Head Start site can.

Until further research data are available, it would seem imprudent to assign either day care or Head Start responsibilities solely to school districts in general. In some communities the school could be the optimal site. But respect for individual differences in children suggests that many kinds of auspices are a safer choice than is delegation of such programs to a single type of auspice.*

3. Age, Duration, and Type of Curriculum

The data in this report do not clearly indicate that either a particular age, or a particular length or type of preschool experience is optimal for all children. Future analyses of these data may assist in these policy areas.

4. Head Start

We see three major implications for Head Start programs which flow from these preschool findings.

a. A large national random sample of new Head Start enrollees and their parents (or whole sites) should be identified so that baseline data** could be collected prior to the program experience. We believe that fears of rejection of such data collection at entry by parents are overblown, that most parents are accustomed to an intake procedure for public services. Without such baselines and an initially tight research design, later evaluations of typical Head Start programs are not possible.

---

* These experiments were carried out under various auspices -- public schools, Head Start Centers, independent agencies and universities. Quality control may be more important than auspice.

** We can make suggestions as to the implementation of this recommendation.

b. A careful review of the extent to which structured curricula are actually being used in Head Start programs should be carried out. Further steps may need to be taken to insure that deliberate learning goals are identified and teaching strategies implemented.

c. The establishment of PCC's as a demonstration program a decade ago, and of Home Start some seven years ago, may have been necessary to try those ideas out. To keep these services separated from "regular" (i.e., center-based) Head Start operations seems no longer justified. The evidence suggests that all Head Starts be encouraged to adopt PCC and Home-Start services (and age groups) as part of their regular programs, to obtain the benefits of earlier and mixed-locus intervention. The separate PCC's and Home Start Centers could be converted into training sites during the period of introduction of their services into Head Start programs, and then could be expanded in scope to become "regular" Head Starts.

Finally, and recognizing the self-serving nature of this recommendation, we think that this study provides a reason why Head Start research funds should be increased, with a greater portion of those funds reserved for investigator-designed rather than contract-designed studies.

PART VII

FUTURE TASKS

As rapidly as resources will permit, we plan to complete analyses of these data to address questions in the following areas:

A. WISC-R Subtest Scores and Patterns

Is there a different pattern among sub-test scores for the control group compared to the program group? Miller has a sample of fourty-four children who took the Stanford-Binet as well as the WISC-R in the spring of 1977. The sample was stratified by original Head Start program or control and by sex in so far as was possible. We plan to investigate the comparability of the two IQ tests.

B. Achievement Tests

We have 3,1[illegible]9 test records for the 1,115 subjects. The tests were given in grades one through twelve, but our results were primarily for grades three, four, five and six. There are 200 records for grades seven and eight and 100 for grade nine. The approximate N's per test are: Gates MacGinitie - 100; California Achievement Test - 200: Comprehensive Test of Basic Skills - 400; Iowa Test of Basic Skills - 400; Wide Range Achievement Test - 400; Metropolitan Achievement Test - 600; and Stanford Acheivement Test - 600. There are grade norms for 1,349 records using primarily national norms. This data has not been analyzed yet because of the very complex transformation problems involved. We will investigate several possible approaches to these analyses.

C. Special Education and Retention

Special education and retention as main effects are established but we do not yet know anything about the pathways to those outcomes and what effects different mental abilities, family structures or sex, for example, have on those outcomes. The effects of "mainstreaming" under PL-94-142 may mask these differences in future studies, making replication with a new sample very difficult.

D. Parent and Youth Interviews

Interview responses are yet to be analyzed in depth. In the area of values, self-esteem and social perception, we have indications of some provocative results. For example, using Youth Interview responses to question #17, "What is the worst trouble you've ever been in?", 20% of the control group compared to 14.9% of the program subjects mentioned police-involved incidents, drugs, thefts, or being expelled or suspended. This difference appears to be the case for both control boys and girls. The program group also appears more likely to have higher job aspirations. Both these results, along with the dropout findings discussed in the analysis section, are only preliminary findings.

The Consortium members have agreed to pool efforts to investigate the extent to which program and controls are known to state Child Welfare Systems. This investigation is in its preliminary phase.

E. Who Benefits?

The basic practical question for which the data will soon be on hand remains a central concern:

What kinds of children
from what kinds of families
benefit most
from what kinds of intervention
at what points - and with what duration - in their early years?

F. Developmental Questions

Aside from the effects of early intervention, what can these data tell us about cognitive and emotional development in children from low income families, during the school years?

PART VIII

APPENDICES

APPENDIX A: STATUS REPORT

APPENDIX A-1

INTRODUCTION

Although this constitutes our final report to OCD on research conducted with their support in FY 77, it does not constitute the final work we intend to accomplish. Some data is still being collected, and we plan to undertake a number of additional analyses. Because the costs for this study were necessarily unpredictable (i.e., the number of subjects that could be found was not predictable, nor the length of time necessary to get school district clearances), the spring meetings of the Consortium were canceled so that the funds could be used for data collection.

As of August 15, 1977, salary moneys were fully expended. A grant from the Hewlett Foundation will permit some continuity until new funds can be obtained. This will permit coding of the additional data which will arrive, and completion of several analyses now in progress. These will be submitted as separate supplements, but are not crucial to the major purpose of this study.

APPENDIX A-2

CONSORTIUM MEETINGS

A preliminary Consortium meeting sponsored by the Social Research Group of George Washington University was held in Washington on July 9, 1975. Here, the interested investigators met to discuss the issues involved in undertaking a pooled longitudinal research effort. Discussion centered on definitions of developmental continuity and on the policy implications of various approaches to the research.

The Education Commission of the States offered to sponsor the subsequent Consortium efforts. At a second meeting, participant researchers agreed as an initial step to make all their original and follow-up data available in detail for the purpose of joint analysis. Additional tasks, such as the production of a position paper and an inventory of available data, were agreed upon as priorities.

A subsequent meeting was held in Washington on October 3, 1975. At that time, Dr. Palmer related the issues of his paper, "Has Compensatory Education Failed?" to the current educational climate among public and private decision makers. Dr. Lazar's draft paper on research issues was also discussed, and plans were made for identification of available data. Dr. Lazar was delegated as chairperson for the study.

A series of regional subcommittee meetings were held late in 1975 and early in 1976. Of primary concern was the selection of instruments to be used by each of the programs involved in the study.

A meeting in March, 1976 in Gainesville firmed up the study design and budget, and specifications for proposals with which to seek support for the longitudinal research. The initial interview forms were approved for pretest.

Following award of a grant from the Office of Child Development, a meeting was held in conjunction with the October, 1976 NERA conference in Ellenville, New York. Discussion centered on the design and use of the various instruments. Issues such as the analysis of SES, attrition problems, and comparability of IQ and achievement scores were brought up. Each project presented a progress report on their data collection efforts. Agreement on variables to be measured was reached, and a tentative production schedule was established.

APPENDIX A-3

PRESENTATIONS GIVEN 1976-1977

Presentations by Consortium members were given at the meetings of the American Psychological Association (September, 1976); Northeastern Educational Research Association (October, 1976); National Association for the Education of Young Children (November, 1976); American Association for the Advancement of Science (March, 1977); American Orthopsychiatric Association (April, 1977); American Educational Research Association (April, 1977); and Office of Child Development Conference (May, 1977). Papers presented at these meetings are listed below, and are attached in Appendix C.

Brown, Bernard, "Methodological Issues in Comparing Early Intervention Programs" (Symposium). Symposium Participants: Irving Lazar, E. Kuno Beller, Susan Gray, Ira Gordon, Victoria Seitz, presented at Northeastern Educational Research Association, Ellenville, New York, October, 1976.

Brown, Bernard, "Sleeper Effects from Early Intervention Programs." Symposium Participants: Cynthia Deutsch, Merle Karnes, Louise Miller, Francis Palmer, Sheldon White, presented at Northeastern Educational Research Association, Ellenville, New York, October, 1976.

*Hubbell, Virginia Ruth, "The Developmental Continuity Consortium Study - Secondary Analysis of Early Intervention Research," presented at the American Association for the Advancement of Science Conference, Denver, March, 1977.

Hubbell, Virginia Ruth, "Differential Effects of Early Childhood Intervention Programs," presented at the American Orthopsychiatric Association Annual Meeting, New York, April 16, 1977.

Karnes, M.B., "The University of Illinois Study of the Differential Effects of Five Preschool Programs", presented at the Annual Meeting of the American Educational Research Association, New York, April, 1977.

Lazar, Irving, Chairman, Symposium: "Early Intervention: How Well Does It Work?" Participants: Francis Palmer, Victoria Seitz, Robert Hess, Discussants: Urie Bronfenbrenner, Marshall Smith. American Psychological Association, Washington, D.C., September, 1976.

Lazar, Irving, with Virginia Ruth Hubbell, Harry Murray, Marilyn Rosche, Jacqueline Royce, "Preliminary Findings of the Developmental Continuity Longitudinal Study," presented at the Office of Child Development "Parents, Children, and Continuity" conference, El Paso, Texas, May 23, 1977.

Miller, Louise B., "Differences in Process and Product in a Four-Program Comparison: The Louisville Study," presented at the Annual Meeting of the American Educational Research Association, New York, April, 1977

Murray, Harry William, "Early Intervention in the Context of Family Characteristics," presented at the American Orthopsychiatric Association Annual Meeting, New York, April 16, 1977.

*Palmer, Francis H., "The Effects of Early Childhood Intervention", presented at the American Association for the Advancement of Science Conference, Denver, March, 1977

Palmer, Francis H., "The Effects of Minimal Early Intervention on Subsequent IQ scores and Reading Achievement," presented at the Office of Child Development "Parents, Children, and Continuity" conference, El Paso, Texas, May 23, 1977.

Palmer, Francis H., "The Effects of Early Childhood Educational Intervention on School Performance," prepared for the President's Commission on Mental Health, July, 1977.

*Seitz, Victoria, "Long-term Effects of Intervention: A Longitudinal Investigation," presented at the American Association for the Advancement of Science Conference, March, 1977.

Weikart, David, "Can Preschool Make a Lasting Difference?" presented at Office of Child Development "Parents, Children, and Continuity" conference, El Paso, Texas, May 23, 1977.

---

* Included in Brown, B. (Ed.), Found: Long Term Gains from Early Intervention, Westview Press, Boulder, 1977 (Copyright, American Association for the Advancement of Science).

APPENDIX A-4

STATUS BY INSTRUMENT

The data collected from six instruments comprise the follow-up data base. Data received at Cornell before July 1, 1977, on four instruments (Parent and Youth Interviews, School Record Forms and WISC-R scores) were coded, keypunched, cleaned and merged with "original" (pretest and earlier posttest) data to form a SPSS Consortium Master File. This Master File is now on computer tape and was used for the analyses reported in this paper. All Woolman data, however, are in a separate file on computer tape. Also, Zigler's earlier pre and post test results are in the Consortium Master File, but the latest posttest data are on computer cards. Data on the other two instruments (Achievement Tests and School Supplements) have been keypunched and cleaned but are not yet included in the Consortium Master File. Table A-1 shows the number of cases received from each project site by instrument as of September 30, 1977. Table A-2 shows the number of cases in the SPSS Master File used for this report. A more complete discussion of the current status of each instrument follows below.

1. Parent and Youth Interviews

For this report, the Consortium Master File includes 871 Parent Interviews and 773 Youth Interviews merged with the pretest and earler posttest data. In addition, the separate Woolman file (which had no earlier posttest data) included fifty-four Parent and ninety-seven Youth Interviews. Since the July 1 deadline, thirty-two additional Parent Interviews and forty additional Youth Interviews have been coded, keypunched, cleaned and are ready to add to the Master File. Interviews continue to arrive at Cornell (primarily from Beller and Gordon sites), but have not yet been coded or keypunched. This group includes an additional thirty Parent and sixty-two Youth Interviews. In addition, 185 Zigler scores will be added to this total. The Zigler site interviewed their subjects as they reached seventh and eighth grade using their own interview. Four or five of their attitudinal questions correspond to questions on the Consortium Youth Interview. These data, which are on computer cards, were sent to Cornell along with PPVT and Achievement scores for those cases. As shown in Table A-1, the total number of Consortium Parent Interviews received by September 30 was 987. The number of Consortium Youth Interviews received was 971 plus 185 Zigler interviews, making a total of 1156. These totals are approximate. It is to be noted that all projects with the exception of Zigler interviewed their subjects in the 1976-77 year.

Other family data yet to be keypunched and added to the file are Beller's retrospective data. These demographic data are to be included in the pretest ("original") part of the Consortium Master File.

2. School Record Form

Data on special education and retention for 1529 cases are included in this report. As discussed in Part II, Methodology, this information came from several sources. The Beller site has sent an additional thirty-six School Records not yet keypunched. As shown in Table A-1, the September 30 total is 1578 School Records.

3. Wecshler Intelligence Scales

Eight project sites reported Weschler scores. Analyses of results for six projects are included in this report. The seventh site, Karnes, has no untreated control group. The Consortium Master File, as of July 1, includes IQ scores for 630 cases. An additional ninety-five from the Woolman site are in a separate file. The fifty-seven WAIS scores received from the Deutsch project have been keypunched and cleaned and are ready to merge with the Master File. Zigler sent 185 Peabody Picture Vocabulary Scores on computer cards which can be included in future analyses. As shown in Table A-1 the number of cases with follow-up post test IQ scores is approximately 979. The Gordon and Beller sites are currently administering WISC-R tests so that the IQ sample is expected to be close to 1000. (See Part II; Data Collection Methodology for additional detail on which sites administered WISC or WISC R tests and how old the subjects were). In order to investigate the comparability of the two IQ tests, the Miller site sent a listing of IQ scores for a sample of forty-four children who were administered the Stanford Binet as well as the WISC-R at this posttest. Karnes is administering both tests to her entire sample.

4. School Supplement

Three project sites chose to collect supplemental school data on attendance, marks, etc. A total of 219 cases were received and keypunched. Approximately 200 cases from the Woolman site also have these supplemental data available. None of these data have been cleaned or analyzed.

5. Achievement Test Scores

As shown in Table A-1, one or more Achievement Test Scores for 1333 subjects have been received and keypunched. These data have not been merged with the Master File, for several reasons. First of all they have not yet been cleaned. In addition, problems merging with the SPSS Consortium Master File have not been worked out. The data for an SPSS file must have exactly the same number of cards for each case. For the achievement test data, each case has a different number of cards depending on how many achievement test scores are reported. The preliminary analysis on which tests were given and at which ages is discussed in Part VII, Future Tasks.

Table A-1: Number of Cases Received by September 30, 1977 for Each Instrument by Project Site

Project Site	Parent Interviews	Youth Interviews	School Record Form	Ach't. Tests	School Record Form	IQ Scores**	Any Data+
Beller	93	93	77	64	64	--	106
Deutsch	53	58	33	21	--	57	66
Gordon	107	76	--	--	--	--	101
Gray	72	69	74	72	74	72	77
Karnes	120	114	107	99	104	112	121
Levenstein	98	75	120	116	--	76	186
Miller	141	141	139	134	--	141	141
Palmer	143	144	209	197	--	132	228
Weikart	106	104	123	96	--	110	123
Woolman	54	97	511	349	200	95	611
Zigler		185	185	185		185	185
Totals	987	1156*	1578	1333	442	979	1945

Note: These totals are approximate. See Appendix A-4 and Part II (Data Methodology) for additional detail.

* Total includes 185 Zigler cases who did not have Consortium Interview.

** Scores include WISC-R scores for Gray, Karnes, Miller, Palmer, Woolman;
WISC scores for Levenstein, Weikart;
WAIS scores for Deutsch;
PPVT scores for Zigler.

+ Any data is defined as number of cases for which data on at least one instrument were reported.

Table A-2: Number of Cases in July 1, 1977 Computer File for Each Instrument by Project Site

Project Site	Parent Interviews	Youth Interviews	School Record Form	WISC	Any Data*
Beller	78	59	41	—	86
Deutsch	47	57	33	—	63
Gordon	96	42	—	—	96
Gray	72	69	74	72	77
Karnes	120	114	107	112	121
Levenstein	84	59	120	76	186
Miller	141	141	139	141	141
Palmer	127	128	198	119	221
Weikart	106	104	123	110	123
Woolman	54	97	509	95	185
Zigler			185		611
Totals	925	870	1529	725	1910**

* Any data is defined as number of cases for which data on at least one instrument were reported for this analysis.

** Total includes 1114 cases in merged file; Woolman and Zigler data are on separate data files.

APPENDIX B: TECHNICAL SUPPLEMENTS

APPENDIX B-1

SAMPLE SELECTION AND ASSIGNMENT

Perhaps the most critical question which can be asked about a report which combines the results of a number of studies is this: Were the subjects initially comparable across studies, or were the studies dealing with different types of children? One approach to answering this question was contained in the body of the report, namely a comparison of the project children on several initial measures (SES, IQ, and mother's education). A second approach is to analyze the actual criteria and procedures used for sample selection and assignment to program and control groups. Since the projects included in this analysis were conducted independently (with the exception of those projects conducted by the same principal investigator), no two had criteria and procedures which were exactly alike. In order to let the reader form his/her own opinion on the comparability across projects, the selection processes of the projects will be summarized in three sections:

- source of subjects
- selection criteria
- procedures for assignment to program or control groups.

Source of Subjects

As seen in Table B-1, most projects recruited their subjects from the general community (usually a fairly well defined neighborhood). Two projects (Gordon and Palmer) recruited their subjects from names on hospital birth records. The Woolman project included children referred to his program by the school district.

Selection Criteria

As shown in Table B-2, a variety of criteria were used in selecting the samples. All projects used geographic and age criteria, and all but Woolman's specified some type of SES criteria. Table B-3 delineates the specific SES criteria used. One should note that the Palmer project was the only one to make a specific attempt to include middle class children as part of its sample. Many projects specified some sort of health criteria, usually designed to eliminate children with organic impairments. Several projects did have special selection criteria which caused their samples to differ from the others in important ways:

- The Palmer project selected only boys.
- The Karnes project selected only children with an IQ of 70 or above (this project has the highest mean pretest IQ).
- The Perry Preschool and Curriculum Demonstration projects selected only children with an IQ below 85. (These two projects have the lowest mean pretest IQ's.)
- Finally, the three New York projects specified that the children must be English speaking (Levenstein relaxed this criterion for later waves), although most projects in other areas would not have found this criterion relevant. Only the Woolman project had a large proportion of Spanish speaking children.

Procedures for Assignment to Program and Control Groups

The procedure used to create the program and control groups is one of the most crucial factors affecting the validity of research findings on program effectiveness. If proper procedures were not employed there is no real way to know whether the program and control groups were comparable to begin with, and, hence, there is no way to accurately measure whether the program had any effect.

A major concern in sample assignment is the question of self-selection. Bernstein et al. state:

> Obviously, if clients themselves determine whether or not to seek treatment, one never knows whether it is the treatment itself which is responsible for observed differences between the experimental and control groups, or whether other variables correlated with the selection of treatment versus control are responsible for the observed effects. (Bernstein et. al., p. 109) (1)

Thus, if a study were designed in such a way that the program children were those whose parents volunteered for the program, while the control group consisted of children whose parents may have decided not to volunteer for the program, the two groups were initially different in an important way. The post-treatment differences between experimental and control groups might be due, not to the program, but rather to the fact that the program children had parents of the type which would volunteer for such a program while the control children did not.

The more generic problem is that, if the assignment proceeds does not closely approximate a randomized assignment to program and control, the results may be biased to an unknown extent. A reviewer who believes that the program was a "good" one would be able to formulate excellent theoretical reasons why the control group was originally "better off" than the program group with the result that the effect of the program was underestimated.

On the other hand, a reviewer who believes that the program was "ineffective" could undoubtedly also create excellent theoretical reasons why the control group was originally "worse off" than the program group and that, hence, the effect of the program was overestimated. Thus, unless the program and control groups are randomly selected from a population which volunteered for the program (a design which is questionable on human relations grounds - unless other services are offered - and which is nearly impossible to implement in our society), doubts can always be raised as to the initial comparability of the two groups. These doubts can be raised even in numerous indicators of initial status (e.g., I.Q., SES) were collected at the outset indicate that the two groups were comparable, since it can always be hypothesized that the "true difference" was not reflected in these measures.

In general, the projects used in this analysis come closer to being experimentally valid. After reviewing their sample selection and assignment procedures, we divided those projects included in the analysis of special education and retention into two groups: those most closely approximating a randomized experiment (Gordon, Gray, Weikart, and Palmer) and those which were more "quasi-experimental" (Levenstein, Miller, and Zigler). Another reviewer could have arrived at a different classification since the differences are not always clear-cut. Therefore, the procedures for assignment to program and control groups will be breifly described for each project so that the reader may make an independent evaluation of their validity.

Kuno Beller's Philadelphia Project

> Each of the four schools opened a nursery program for fifteen four-year-old children. Applicants were recruited through notes to parents of all pupils attending each of the four schools which announced the opening of such a program . . .
>
> Fifty-six of the original children graduated to kindergarten in the same four public schools in which they had attended nursery school. Group II consisted of fifty-three five year olds who entered the same kindergarten classes as the children of Group I however, without prior nursery experience. These children were selected from a larger group to approximate age, sex distribution,

and ethnic background of the children in Group I. The majority of children in Group I and II graduated from kindergarten to first grade classrooms in the four same schools in which the original program started. All children from Group I and II were assigned to first grade classrooms in each of the four schools in such a way that an equal proportion of children from Groups I and II would have the same teachers. This was done to reduce differential effect of the educational experiences due to differences between classrooms and teachers. From the first grade classrooms in which fifty-eight children of Group I and II were enrolled, a third group of children was selected who had no prior preschool experience. Again these children were selected to be comparable to the age, sex distribution, and ethnic background of children in Groups I and II.

Martin and Cynthia Deutsch's Institute for Developmental Studies Program

As the program became established, an increasing number of parents volunteered their children for it. However, particularly in the early stages of the program's operation, it was also necessary to conduct active recruiting in the community. Staff members obtained the names of children from a variety of sources: from church groups, from response to posters, from word-of-mouth information, from other children, from teachers and principals, etc. Minimal demands were then made on the parent(s), who were asked to respond to a short interview given in their home and to bring their child to the school for a short observational period.

A final sample was chosen from the group that had been gathered through these recruitment procedures.

Approximately one third of the children selected by the above criteria could not be handled by the small number of Institute classrooms, and these children, chosen randomly from the total N, constituted a control group. Since this group was equivalent to the experimental group in parental motivation and desire for the child to be part of an enrichment program, it was designated the control group for the factor of self selection, and is is referred to subsequently in this report as the Css group. The experimental group was given enriched schooling from prekindergarten through the third grade, while the Css group first en-

> countered formal schooling in the regular kindergartens (or, in some cases, first grades) of the New York City public school system. (Although most Css children began school at kindergarten, this was not stipulated as a requirement when the group was originally formed.)
>
> In order to provide further controls for the IDS program, two additional control groups were constituted, one at the beginning of each of the next two successive years. The first was the "kindergarten control" or Ck group. This was composed of children who had no prekindergarten training and who entered the kindergartens of the same public schools that the E and Css groups attended. When the Ck group was selected, of course, the E and Css groups were also entering kindergarten. Apart from the time of their selection, the children in this group differed from those in the Css group primarily in that they were not recruited nor had they volunteered for the Institute's program. The second additional control group was constituted at the beginning of the following school year, and included children who entered the same public schools as the other groups at the first grade level. They had no prekindergarten and no kindergarten experience.
>
> In the later years of the program, Head Start was in full swing, and, to obtain a sufficient number of children for the Ck group, it was necessary to include children who had been in the Head Start program. Their inclusion, of course, would have the effect of minimizing E vs. Ck differences, and thus was an experimentally conservative choice.

Gordon's Parent Education/Backyard Learning Center Project

The procedures used by the Gordon project to assign children to program or control groups were quite complex and will only be summarized here. Three waves of children entered the project. For the first wave:

> Assignment to experimental or control groups was based on randomization of geographic area to avoid contamination. Towns and their surrounding areas were randomly assigned as experimental or control towns then they were randomly assigned as Negro or Caucasian towns. A given town became experimental for one race and control for the other. Gainesville, which has a relatively large population, contains control and experimental subjects of the same race but in different areas . . . (2)

Mothers of children in program areas were invited to participate

> Half the control population (C_1) were also contacted and invited to participate in a program for helping the University learn how babies grow. They were informed that they would be visited about once a month and that some testing of the baby's development would be made periodically. The remainder of the control population (C_2) was not solicited for active participation until the baby reached his 12th month birthday, at which time a parent educator visited the family and invited the mother to bring the baby in for testing. (3)

For the second wave:

> . . . three new groups, (E_2, C_3, and C_4) selected in fashion as the original population . . ., and randomly assigned to series stimulation, "other" stimuation both considered experimental for the current analysis and a new control group, were started in July 1967. They were drawn from all the eligible babies born in the hospital between May 1, 1967 and October 31, 1967.(4)

At the start of both the second and third years of the project, children from these two waves were randomly reassigned to experimental or control status for that year. At the start of the third year of the project, a new wave of children were added who were of the same age as the original children (i.e., now two years old). (5)

This wave was randomly assigned to experimental or control status for the final year of the project. Thus, all parents volunteered for a program to study young children and all children had the possibility of being assigned to experimental or control status for at least one year of the project.

Susan Gray's Early Training Project

Sixty-one low-income children were randomly assigned into three groups: two groups which received the program (beginning at different ages) and one control group. The two program groups are treated jointly in this analysis. Since program and control groups were randomly assigned from the same population (a group of children whose parents volunteered for the program), there is no self-selection factor contaminating a comparison of program and control groups (both volunteered with the possibility of recieving the program). A second control group was established in another city (6). This group is included in most of analyses in this report, but is excluded from the analysis of special education and retention because of the self-selection factor (since this group was recruited specifically as a control group) and because this group was in a different school system (which would contaminate the results on special education and grade failure because of differences in school policy).

Susan Gray's Family Oriented Home Visitor Program

The children were randomly assigned to one of three groups: Extensive Home Visiting (EHV), Materials Only Home Visiting (MOHV) and Control. There were three waves of children in the project.

Merle Karnes' Research and Development Program on Preschool Children

Two waves of children entered the Karnes project. The first wave was assigned to one of three curricula: GOAL, Bereiter-Engelmann, or traditional. The second wave was assigned to one to one of three curricula: Bereiter-Engelmann, Montessori, or Community Integrated. There were no untreated controls. For each wave the assignment procedure was as follows:

> The 1960 Stanford-Binet Intelligence Scale was administered to eligible children who were then stratified on the basis of their intelligence quotients into three groups: IQ scores 100 and above, 90 through 99, and 70 through 89.4. The children were assigned to class units (N=15) in

> which one-third of each class consisted of children who had scored in the "high" IQ range; one-third, in the "middle" range; one-third, the "low" range. Mean intelligence quotients were then computed for the three strata and for each class unit. These means were evaluated for comparability between class units as a whole and for strata between classes. These strata insured a balance range of intelligence scores in each class unit and provided an opportunity to evaluate the effectiveness of the various programs on children from different ability groups. The mean IQ (approximately 95) of children placed in classes is, of course, higher than the mean of children screened.
>
> Class units were examined to assure comparability of sex and race. When necessary, substitutions were made between classes to maintain an approximate ratio of 67% Negro children and 33% Caucasian children and a ratio of approximately 50% male and 50% female children. Finally, each class unit was randomly assigned to a particular intervention program. (7)

Phyllis Levenstein's Mother-Child Home Program (8)

The procedures used by the Levenstein project to assign children to program or control status were "quasi-experimental." Three cohorts of children are used in the analysis of special education and grade failure.

The first cohort consisted of children from three housing projects; one project was randomly designated as a program project, while the other two were control projects. The children in the control projects were recruited for testing only; however, one of the latter groups were given the program during the second year (i.e., when the children were three years old), and, hence, is treated as a program group. The second cohort consisted entirely of program children, while the third contained two groups - one group recruited to receive the program and a control group recruited to receive the program materials only (i.e., no instruction was given). An "after-only" control group was recruited in the first grade; however, the extent to which self-selection would make this group not comparable to the program groups is unknown. It would seem on the basis of the selection procedures described that the other two control groups would be more likely to be comparable to the program groups (since their parents did volunteer them for participation of some sort in their preschool years, if only testing or receiving materials). More importantly, since the group was drawn from a first grade population, children who were assigned to special education prior to first grade or in the early stage of first grade would be underepresented (i.e., the sample was drawn from regular first grade classrooms). Therefore, this group is not included in the special education or grade failure analyses, but is included in I.Q. test analyses.

Louise Miller's Experimental Variation of Head Start Curricula

The program groups consisted of 214 children enrolled in Head Start classes in Louisville, Ky. The control group consisted of thirty-four children from the same neighborhoods - twenty-one of whom were in the Head Start waiting list are unknown. The program and control groups were not selected from the same population; therefore, there could be an initial difference due to the self-selection factor (at least in the case of the thirteen children who were not on the waiting list). The control group contained significantly more white children and significantly more children with both parents in the household. Thus, there is reason to believe that the program and control groups are not totally comparable.

The procedures used to assign program children to one of the four variations in curricula were as follows:

> The four experimental replications were placed in four "target areas" of the city....Thus it was important to assess sample characteristics and to balance classroom facilities across programs. Since the experiment was being conducted as a part of the regular Head Start Program, it was also desirable to provide both experimental and non-experimental classes in the same schools.
>
> The four target areas were designated California, Jackson, Park-DuValle and Russell. Since the 4-program comparison could be replicated in only two areas with the two Montessori teachers, the two largest area, Park DuValle and Russell, were selected for this purpose. All four areas contained replications of the 3 program comparison. Classroom facilities in the Russell area were in general inadequate. All four programs were located outside of school buildings, three in churches and one in a small and very old portable. In the other three areas, facilities were in satisfactory school classrooms.
>
> To some extent the distribution of classes constituting the various programs into different geographical areas insured that the combined program samples would consist of similar subjects. But it was also desirable that children who attended experimental classes would constitute a random sample of those who registered for Head Start.

In the strict sense, "random" assignment of subjects would be accomplished by obtaining the names of all children eligible for Head Start and, assigning each child to one of the fourteen classes or to a control group by using a table of random numbers. This would not have been possible with a sample of approximately 250, since it would have forced many children to cross the city to attend schools outside their neighborhoods.

> It was possible, however, to arrange for assignment of registrants in each school on a random basis. All schools contained at least two Head Start classes. In one school, both classes were experimental; however, in the remaining schools both experimental and non-experimental classes were available for distribution of subjects. Registration forms were filled out on the same day in all schools. The parents' signature on the form gave permission for children to be placed in experimental classes should they happen to be selected.
>
> When all forms were turned in, the forms were divided on the basis of sex to insure a balance in each class. These piles were then shuffled and distributed into classes, experimental or non-experimental. (9)

Thus, the procedures used by Miller for assignment to different experimental groups were quite rigorous since the comparison of curricula was the primary purpose of the project (rather than the comparison of program children with untreated control children).

Francis Palmer's Harlem Training Project (10)

The program children were selected from the children born between August 1 and October 31, 1964, in the Harlem and Syndenham Hospitals. The controls were born in November-December, 1964, in the same hospitals. The program children were recruited for the program, while the controls were recruited specifically as controls. However, in recruiting the controls, the benefits from a total of four and one-half weeks of testing were stressed so that, in effect, the parents were volunteering for a program though of much shorter duration. Although the two groups were not technically drawn from the same population, there is little reason to suspect a self-selection bias.

David Weikart's Perry Preschool Project (11)

Five waves of children (a total of 123) volunteered for the program.

> The study took place in Ypsilanti, Michigan, a city of about 30,000, located about 30 miles west of Detroit. Children came from families in the Perry school attendence area. This neighborhood was selected because the school had a history of low academic achievement. Each September, the names of all families with 3-year-old children were drawn from the school census; the socioeconomic status (SES) of these families was determined from the parents' education, occupation of the head of the household, and household density (rooms per person). If the SES score of the family was below a specified level, the 3 year old was given the Stanford-Binet Intelligence Test. Children with low Binet scores (50 to 80), but no evidence of organic impairment, became a part of the study sample. Reflecting the ethnic makeup of the neighborhood, all the children in the sample were black. Thus, the study sample consisted of young black children of low measured IQ from families of low SES.
>
> Each year children were assigned to preschool or no preschool by the following procedures:
>
> 1) Rank the children by their Binet scores; make an odd/even sort into a preschool group and a no preschool group.
>
> 2) Exchange children with similar scores to equate the sex ratio and average SES score of each group.
>
> 3) In later waves, assign younger siblings to preschool if their older siblings attended preschool, to no preschool if their older siblings did not attend preschool. (This was done to keep preschool from having an indirect effect on the siblings who did not attend preschool.)
>
> 4) Exchange children with similar scores when a child assigned to preschool cannot attend due to lack of transportation or a working mother. (No funds were available for transportation or full-day care, and special arrangements could not always be made.)

David Weikart's Curriculum Demonstration Project

Data included in this report are for the middle two waves of children. Children with a brother or sister already in the project were assigned to the same program as their siblings. The remaining children in each wave were divided into three equal-sized groups matched on ethnicity, sex and age to the extent possible. Twins were assigned to the same groups. The three groups were then randomly assigned to program types.

David Weikart's Carnegie Infant Program

> The experimental design specified that sample children would be randomly assigned to three different treatment groups--two groups (experimental and contrast) to be seen by either teachers or community representatives and the third group (control) to receive only the same testing and data collection as the other groups. In order to avoid the possibility of biasing the original sample by accepting only families who were willing to be assigned to a "no-treatment control group" or alternatively by losing families following assignment to a "no-treatment group," none of the three levels of participation was described as "no-treatment."
>
> After a family's eligibility had been determined, the teacher explained to parents that project staff were interested in the different ways children learn by playing; then she asked the parents if a staff member could periodically bring toys to the home to be used by the mother and the staff member for playing with the baby. This described, equally well, both the home teaching sessions and the administration of the Bayley Scales of Infant Development, so that regardless of which treatment the family received, the same initial description was accurate. All parents were told that the project would last sixteen months, that the mother must be present at each session with the child, that no session would be held without her, and that she could choose the times for the appointments.
>
> After a sufficient number of families had agreed to participate, members of the research staff assigned them to treatment groups according to a table of random numbers. During this random assignment the research staff did not know which names belonged with each treatment group, since they were numerically assigned Four sets of two children each were assigned jointly; two sets because the children were twins and two sets because the mothers were sisters.

> Finally, all families assigned to the experimental treatments (experimental and contrast groups) were again approached and asked if they would allow a staff member to bring toys on a weekly basis for use by the mother and staff member together for playing with the baby. Only one family could not comply with this request and had to be dropped from the sample. Families assigned to the control group (those receiving testing and data collection only) were not approached with this request. However, since they were never told that the regular visits to their homes by research staff did not constitute treatment, control group families felt that they were participating in an infant education program.

The contrast group is not included in the current paper. A fourth group, consisting of children in the original sample who did not become part of any of the three groups but who were recontacted in first grade, is included in some analyses.

Myron Woolman's Micro-Social Learning System

The Woolman project consisted of one experimental group and three control groups. The experimental group was selected by the Vineland school district. The group consisted primarily of the children of migrant families, the majority of whom were on welfare. About 10% of the children were of higher socioeconomic status (These were children whose parents had requested that they be allowed to enter the project).

The "initial control" group was selected by the state of New Jersey and consisted of children who had had other preschool experience. Sample data indicates that the control children scored significantly higher on pretest IQ. However, pending a more complete description by the state of the exact selection procedure, this group has been excluded from analysis.

The "parallel control" group consists of a random sample of children currently (in 1977) in the same classrooms as the experimental children. This group was considered inappropriate for the special education-grade failure analysis. (Obviously, there would be no difference in those in special education classes since children were selected from the same classes. Similarly, the question of grade failure would be confounded since the children selected from the classrooms of those who had failed would consist primarily of children who had entered school in a different year). The group is, however, used in the IQ score analysis.

The third control group, known as the "baseline controls," consists of a random sample of the general school district population which entered the school in the year preceding the experimental group. This group was the one selected for the special education and grade failure analysis and is further described in that section.

Edward Zigler's New Haven Follow Through Study

Two waves of children were used in the study. In the first wave the program groups consisted of children recruited for Follow Through in several low-income areas in New Haven. The control group consisted of "all of the economically disadvantaged children in one classroom from each of three schools located in similar low-income areas." For the second wave of children, the control group consisted of children drawn randomly from the same schools from which the controls for the first wave had been drawn. Since the program and control children were not drawn from a common pool of children whose parents had volunteered them for Follow Through, the extent of the self-selection bias is unknown. It is likely to be minimized by the fact that the controls were drawn from different schools than the program children (meaning that the controls did not consist of children whose parents had decided not to participate in the program). (12)

Table B-1

Primary Source of Subjects by Project

Project	Source
Beller	general community
Deutsch	general community
Gordon	hospital records
Gray-ETP	general community
Gray-FOHV	general community
Karnes	general community
Levenstein	general community and public housing projects
Miller	Head Start lists*
Miller	Head Start lists*
Palmer	hospital records
Weikart-Perry	school district census
Weikart-CD	school district census
Weikart-Carnegie	school district census
Woolman	general community and school referals
Zigler	general community

---

* Part of the control group was drawn from the general community rather than from Head Start lists.

Table B-2: Sample Selection Criteria Used by Projects

	Location	Age	SES	Sex	I.Q.	Language	Race	Physical or Mental Health
Beller	Philadelphia	4,5,6	See B-3					no serious mental or physical handicaps
Deutsch	Harlem	4,5,6,7,8	See B-3			English	Black	generally good physical condition, no serious emotional disturbance or behavior problem
Gordon	Northern Florida	0,1,2	See B-3					single birth, no breach or Caesarian delivery, no complications, no evidence of mental retardation o mother illness
Grey ETP	Murfreesboro or Columbia, Tenn.	4,5	See B-3				Black	
Grey PCHV	Nashville, Tenn.	0,1	See B-3					
Karnes	Champaign -Urbana, Illinois	4	See B-3		70 or above*			
Levenstein	Long Island	2,3	See B-3			early waves were English only		
Miller	Louisville	4	See B-3					
Palmer	Harlem	2,3	See B-3	male		English	Black	5 lbs. or more birthweight, no mothers with histor of narcotics addiction or syphillis, no twins
Weikart-Perry	Perry School Dist., Ypsilanti, Mich.	3,4	See B-3		less than 85		Black	no "organic involvement" in retardation
Weikart-CD	Ypsilanti, Mich.	4	See B-3		less than 85		Black	no "organic involvement" in retardation
Weikart-Carnegie	Ypsilanti, Mich.	0,1	See B-3					
Woolman	Vineland, N.J.	4,5	See B-3					
Zigler	New Haven	5	See B-3					

* For longitudinal groups (one pre-post group not included in the current follow-up was composed of retarded youngsters).

Table B-3: SES Selection Criteria

	Income	Education	Occupation	SES	Housing
Beller			mother working		
Deutsch				.low by institute scale	
Gordon	"indigent" (1st group OEO criteria (2nd)				
Gray ETP	below poverty level	ninth grade or below	skilled or unskilled		poor housing
Gray POHV			unskilled, semi-skilled, welfare; also mother not working or full-time		project, subsidized, or deteriorating
Karnes		educationally deprived		economically deprived	
Levenstein		high school or less	skilled or less		eligible for public housing
Miller	Head Start criteria				
Palmer				half lower class, half middle on the Hollingshead ISP	
Weikart-Perry				Cultural depriv-ation below eleven	
Weikart-CD				"	
Weikart-Carnegie				"	
Zigler		high school or less	semiskilled, unskilled, unemployed		low income housing

APPENDIX B-2

FOLLOW-UP DATA COLLECTION AND DATA ANALYSIS METHODS

An overview of the methodology for the collection and analysis of the follow-up data was described in the body of this report under Part II, Data Collection Methodology. Technical detail on the design of the instruments, data collection and coordination methods (including interviewer characteristics), and data analysis methods (including response rate data) are discussed in this appendix. Copies of the instruments and instructions are found in Appendix B-6.

I. Parent and Youth Interviews

A. Data Collection Methods

After pretesting in the summer of 1976, the final drafts of the Parent and Youth Interviews were circulated to Consortium members for comments the first week in September. A training session for the Interview Field Supervisors was held in Philadelphia on September 17. The supervisors from seven project sites attended and discussed shared concerns.* After incorporating the suggestions of various sites, the final Parent and Youth Interviews were printed and distributed the beginning of October. Video-taped training sessions were prepared at Cornell and distributed to all sites for their continuing use in training interviewers as they were hired. Interview Instructions and other training aids such as "Ethics of Interviewing" were sent with the interview materials.

The Consortium members met during the Northeastern Educational Research Association meeting (NERA) at the end of October to discuss any problems with the interviews as well as to make decisions on the data to be collected on achievement tests and school records. The Interview Instructions were then up-dated throughout the fall as recording ambiguities were found. No changes, however, were made in the wording of the actual questions. Interviewers at each site brought their problems to the site supervisor who then communicated frequently with Cornell staff by telephone and letter. Clarification for Interview Instructions and status reports on data collection were sent to Consortium members regularly and the final revised Interview Instructions were distributed in early January.

---

* The site supervisors were experienced professionals or graduate students and were of various ethnic groups.

The interviewing, which had begun in October, 1976, for some sites, was completed by most sites before July 1, 1977. Two sites which were delayed for a variety of reasons will complete their interviewing and testing this fall.

1. Informed Consent

Either consent/permission forms were signed at the time of the Parent Interview, or an authorization was signed and sent through the mail before the interview, or a verbal request was followed by a letter. (All projects except Woolman obtained written consent before testing or interviewing children. However, permission for testing was obtained from all schools.) Only twenty-two parents refused to participate.

2. Payments to Respondents

The amount of money paid to respondents was an individual project site decision. All sites except one allocated money for respondents. Typically, the sites allocated $10 for the interviewing. One project paid the youth the $10, and did not pay the parent. Two sites paid the parent $10, and did not pay the youngster. Two projects paid the parent and the youngster $5 each, and one project paid the parent $6 and the youth $4. Two sites allocated less money -- one paid the parent $3, and the other site paid the parent $5 and did not pay the youngster. One project paid an additional $5 to the parent (above teh $10 interview payment) for the WISC administration.

3. Characteristics of the Interviewers*

The ten project sites for which we have received this information employed a total of fifty interviewers. The Parent Interviews were conducted primarily by female interviewers (80.7%); male interviewers conducted the remaining 19.3%. A total of 75.6% of the Parent Interviews were done by white interviewers, 23.2% were conducted by Black interviewers, and 1.2% by Cuban or Puerto Rican interviewers.

A total of 51.2% of the youths were interviewed by male interviewers and 48.8% were interviewed by female interviewers. White interviewers conducted 65.6% of the Youth Interviews and Black interviewers interviewed 34.4% of the youths. Six projects reported data on the educational background characteristics for thier thirty-two interviewers. A total of 34.4% of the interviewers had a BS degree or were presently undergraduate students; 53.1% had a Masters Degree (usually in Education or Psychology) or were currently graduate students; 3.1% were from the community without advanced degrees but were previous project staff members; and 9.3% were from the community who were hired for this follow-up study (e.g., a former high school basketball star). A total of one-third of the interviewers were staff members previous to this follow-up study.

---

* Data from ten sites.

4. Characteristics of the Interview

a) Place of the Interview: Most of the Parent Interviews took place in the parents' homes (65.7%); some parents were interviewed in the project center (9.5%); a few were interviewed by phone (.3%); and the remainder were interviewed in other placed such as a college or Board of Education, and occasionally at mother's place of work (24.5%).

b) Interest of Parent in Interview: The interviewers rated 71.3% of the parents as highly or strongly involved and interested in the interview; 24.6% of the parents were rated as neutral; and 4.1% were rated as not involved or interested at all.

c) Length of Interviews: The Parent Interview typically lasted about thirty minutes. The range was from ten to ninety minutes. The Youth Interview typically lasted fifteen to twenty minutes, with a range of five to ninety minutes.

B. Data Analysis Methodology

1. Office Coding Methods and Reliability

Thirteen Office Codes for the open-ended questions on the interviews were developed at Cornell after about 350 interviews were received from the field. The Hollingshead Occupational Scale (1957) was used to code the occupations of household members, the job aspirations expressed by the youth, and the parents' job aspirations for the young person. The occupation of household members was coded in the field by supervisors and carefully monitored for consistency at Cornell. When the first five interviews were received from each site, the Hollingshead codes were recorded at Cornell and any discrepancies were discussed by phone or letter with the site supervisor. This process was repeated until the necessary reliability was achieved for each site. After that point, the occupation codes were checked for consistency when the other office code reliability procedures were done. The major share of the office coding was done at Cornell in the spring by five coders and a coding supervisor. In July, a 5% random sample of all Parent and Youth Interviews was done by two independent coders. The coding reliability was computed for the total of thirteen codes using the percentage of agreement method. The resulting figures were 98% agreement on the total of 2,687 items for the thirty-nine Parent and forty-three Youth Interviews sampled. For the individual office codes, the exact agreement ranged from 90 to 100% on every Office Code. When the percentage of agreement was adjusted for agreement within a category (for example, cognitive aspects of the program), the percentage of agreement for the thirteen Office Codes rose to a range of 95 to 100%.

In order to minimize keypunching problems and later data cleaning problems, careful editing of every interview was done before sending it to the keypuncher. Each interview was checked for correct skip patterns, valid codes, legibility of entries, deletion of names, and empty boxes.

Response rate computation for Interviews is discussed below in Section V.

II. Wechsler Intelligence Scale

The Wechsler scores (WISC, WISC-R and WAIS) were sent to Cornell in a common format on coding sheets. A total of 725 Wechsler scores were received from seven project sites in time to be used in this report. An additional fifty-six scores for the WAIS will be analyzed for the sites that will be administering Wechslers over the summer. The Zigler project sent 185 PPVT scores.

Six project sites provided information on WISC tester characteristics. They hired twenty WISC testers; some were also interviewers, but others were hired especially to administer the Wechslers. Ninety percent of the testers were graduate students or people with advanced degrees in psychology or education. Response rate computations are discussed below in Section V.

III. School Record Form (Face Sheet and Supplement)

For the October NERA meeting, Cornell staff developed an instrument for collecting school characteristics and student performance (marks, attendance, discipline). After Consortium members expressed concern about the time and money involved in collecting this data for every grade, the decision was made to develop a short form to be used by all sites and an optional long form. The School Record Form (Face Sheet, short form) provided data on the most critical variables: current status in school, placement in special education, retention, or skipped experience. In addition, school identification by code number, location and type were collected so that school variables such as funding level, staffing and ethnicity, for example, could be investigated through secondary sources at a later date. Each site agreed to provide school and achievement test data on four years (the latest available year and three other specified years) rather than every year.

The School Record Supplement (optional long form) was designed to provide data on attendance; marks in reading, science and math; curriculum assignment (regular, college prep, voc-tech); psychological evaluation; disciplinary citations; and total number of schools attended.

The School Record (Face Sheet) was mailed to the sites in January along with the achievement test forms and Instructions. The School Record Supplement was mailed with instructions in February.

As of August 15, data for 1,529 School Records (Face Sheets) had been received at Cornell. Three project sites obtained the Supplement data for a total of 219 subjects. Additional Supplement data will be available on a sample of subjects from a fourth site later this summer.

Recorder characteristics and reliability will be discussed below in Section IV.

IV. Achievement Test Scores

Because of the lack of funds, it was not possible for Consortium field staff to administer individual achievement tests. The decision was made to obtain the achievement scores from tests administered by each school system. The Consortium members, after permission was granted by parents and schools, sent field staff to the local schools to record the data from the students' school files. However, the hundreds of schools attended by the students over many years administered a wide variety of tests, with a large variety of test batteries, form and norms, and with scores reported by a variety of methods (raw scores, percentiles, stanines, grade/age equivalents, and standard scores). A universal form was designed by Cornell staff and sent to project sites with preliminary instructions in January. As of August 15, one or more Achievement Test scores for 1,305 subjects had been received at Cornell.

1. Characteristics of School Record and Achievement Test Recorders

Six project sites (Beller, Deutsch, Gray, Levenstein, Miller, Karnes) provided data on who recorded the data from the school files. The interviewers on most field staffs also recorded the data from the school files. They were predominately undergraduate or graduate students or had college degrees. One project (Levenstein), however, arranged for classroom teachers to record the school record data on forms provided by the Levenstein project.

2. Reliability of Coding the School Record and Achievement Test Scores

For two project sites (Woolman and Gray) a photocopy of each of the student's actual school records were sent to Cornell to be transcribed independently by coders at Cornell. For the Woolman site, seventy-one achievement test records were transcribed at Cornell from the original school records to compare with the recording done by field recorders. Using the percentage of agreement method, the resulting agreement between Coder I (in the field) and Coder II (at Cornell) was 98%. The highest category of disagreement was on the recording of battery or level of test administered. For the Gray site, a random sample of seven complete school folders (10%) were sent to Cornell. The agreement was again high between Coder I and Coder II. The percentage of agreement was 94% for the School Record (Face Sheet), 93% for the Supplement and 98% for the achievement tests, when adjusted for the discrepancies between information not contained in the student file but available to field workers who could obtain additional information from the school system. Although it was not possible to check the reliability of the other field recorders, it can be reasonably assumed that the agreement would be similar given the same high standards shown by the project sites in the other data collection processes. Efforts were made to keep the recorders blind as to the treatment vs. control group status of the subjects.

NOTE ON RAW DATA FORMS:

All data reporting and interview forms were printed on self-carbon multisets, so that all original recordings produced two copies. One of these was sent to Cornell, the other retained by the investigator, thus eliminating the need for transcription and chance of error in transmitting the raw data.

V. Computation of Response Rates

In order to calculate response rates for the follow-up sample, an attrition-disposition work sheet was sent to each project site (see Appendix B-6). Six sites (Gray, Karnes, Levenstein, Miller, Weikart and Woolman) sent the data to Cornell in time to be included in this report. The final disposition of each ID case number in the six sites was assigned to one of the following categories.

(C) Completed = Code 1

(R) Refused to give permission, refused to be interviewed = Code 6

(U) Located but unable to test because

Moved = Code 4

Unable to test, keep appt., etc. = Code 5

Terminated, unable to complete test = Code 7

(L) Lost, unable to trace on records = Code 3

(A) Attempted to locate and test: (U) + (L) = Codes 3 + 4 + 5 + 7

(OS) Out of sample

Wave not scheduled to test at this follow-up = Code 2

Dropped from sample at previous follow-up = Code 8

No Data, unknown disposition = Code 9

Response rates are defined as the number of final dispositions in a given category divided by the number of cases in the total sample or subsample. The three response rates computed for completion and refusal categories are shown below:

Response Rate_1 : Disposition divided by total possible cases

$$\text{e.g. } CR_1 = \frac{C}{C + R + A + OS}$$

Response Rate_2 : Disposition divided by attempts (excludes out of samples)

$$\text{e.g. } CR_2 = \frac{C}{C + R + A}$$

Response Rate_3 : Disposition divided by actual cases contacted (excludes lost)

$$\text{e.g. } CR_3 = \frac{C}{C + R + U}$$

The Response Rate Computations and Final Dispositions by Instrument for the six project sites are shown in Table B-4.

Table B-4: Response Rate Computations and Final Dispositions for Six Sites

Disposition	Rate Equation	Parent Interview		Youth Interview		School Record		Ach't Test		WISC	
		N	%	N	%	N	%	N	%	N	%
(1) Completions	$CR_1 = \frac{C}{C + R + A + OS}$	590	42.4	600	43.1	1072	77.0	860	61.7	606	43.5
(2) Completions	$CR_2 = \frac{C}{C + R + A}$	590	74.4	600	78.0	1072	87.4	860	83.8	606	78.2
(3) Completions	$CR_3 = \frac{C}{C + R + U}$	590	82.3	600	86.7	1072	93.1	860	91.0	606	80.9
(4) Refusals	$RR_1 = \frac{R}{C + R + A + OS}$	22	1.6	18	1.3	16	1.1	17	1.2	18	1.3
(5) Refusals	$RR_2 = \frac{R}{C + R + A}$	22	2.8	18	2.3	16	1.3	17	1.7	18	2.3
(6) Refusals	$RR_3 = \frac{R}{C + R + U}$	22	3.1	18	2.6	16	1.4	17	1.8	18	2.4
(7) Attempts	$AR = \frac{A + C + R}{C + R + A + OS}$	793	57.0	769	55.2	1227	88.1	1026	73.7	775	55.6
(8) Not Attempted (Out of Sample)	$OS = \frac{OS}{C + R + A + OS}$	600	43.0	624	44.8	166	11.9	367	26.3	618	44.4

* Total includes 13 Parent Interviews and 16 Youth Interviews received late and not included in other data analyses for this report.

As shown in Table B-4, out of the 1393 original subjects in the six sites, 590 Parent Interviews were completed, resulting in a completion rate (1) of 42.4%. The rate of completion (2) based on the number of parents located and attempted to test is 74.4%. The completion rate (3) based on the parents actually contacted for this follow-up is 82.3%. The refusal rate (6) for the parents actually contacted is 3.1%. The refusal rate (5) based on the total number of parents attempted to locate is 2.8%. The refusal rate (4) based on all possible parents in the original sample is 1.6%. The total number of parent interviews attempted divided by the total number in the original population (7) yields a result of 57.0%. This result may be considered a sampling fraction or a measure of the effort to locate the parents which was dependent on time, money, luck, etc. In contrast, the sample not attempted (8) or out of sample remainder is 43.0%. As discussed in more detail in the technical supplement on attrition, the final samples are generally representative of the original samples in terms of differential rates of program and controls found, pretest Stanford-Binet IQ scores, Hollingshead ISP and mother's education.

The results are similar for the Youth Interview except that the refusal rate is lower. The refusal figure of eighteen for the Youth Interview includes both youths who refused and parents who refused to have their child interviewed.

The refusal rates for the Youth Interview are 1.3% based on total original sample (6); 2.3% based on number located and attempted to test (7); and 2.6% based on subjects actually contacted (8). The completion rates for the Youth Interviews are: 43.1% based on total subjects in original sample (1); 78.0% based on located subjects (2); and 86.7% based on number of subjects actually contacted for this follow-up (3). The percentage attempted is 55.2% and the out of sample percentage is 44.8%.

The Wechsler IQ response rates are similar to the Youth Interview figures. The School Record and Achievement have a much higher completion and attempted rate primarily because the Woolman site had access to all school records and the Weikart site sent data on all school records as of fourth grade.

In summary, the refusal percentages of 2-3% are acceptably low and the completion percentages appear satisfactory given the financial and time constraints.

APPENDIX B-3

ATTRITION

One of the most serious threats to any longitudinal study is the problem of attrition -- the fact that, over time, some of the subjects drop out of a study (due to moving, death or a myriad of other reasons).

> "Differential mortality effects stem from differential subject and program losses from social experiments. If the loss of observational units is different for the treatment and control groups, and the differences cannot be attributed to chance, then external validity is threatened." [Bernstein et al., p. 131]

The analysis of whether attrition has caused biases in the final sample will be directed towards answering three specific questions:

1) Were there different rates of attrition for program and control groups?
2) Were the final samples different on some important characteristic from those who dropped out?
3) Were there any instances of differential attrition (i.e. do significant differences on the characteristics emerge when the sample is divided into four groups: program final sample, program dropouts, control final sample, and control dropouts)?

For purposes of this report, attrition is defined as the failure to report information for a particular child on a particular instrument. This definition is applied because the different methods for collecting information occasionally resulted in considerable divergence among the samples receiving different instruments. For example, collection of Youth Interview and WISC-R data required actual contact with the child, while collection of the School Record Form data required, instead, actual contact with the child's school records -- a feat which could be either more or less difficult than actual contact with the child, depending on the cooperation and organization of the school district.

Attrition in the Current Follow-Up

There are five measures of attrition in the current follow-up:

-- whether a Parent Interview was reported.

-- whether a Youth Interview was reported.

-- whether a School Record Form was reported.

-- whether a WISC-R was reported.

-- whether any of the above was reported (referred to hereafter as "general attrition").

Analyses are presented for each of these five measures of attrition. However, the written discussion will concentrate mainly on the final measure.

The percentages of children found (on any instrument) are given in Table B-5. Since all projects but one (Deutsch) found roughly the same percentage of control children as they did of program children, there appears to be no indication of different rates of attrition between program and control groups. Thus, for all but one project the answer to the first question is negative. However, the percentage found does vary widely among projects. Tables B-6 through B-9 present similar analyses for the individual instruments.

To answer the second and third questions, three characteristics were selected: pretest Stanford-Binet IQ scores (or other IQ scores as noted in the tables), Hollingshead Index of Social Position, and mother's education (in grades completed). Two by two ANOVA's were performed as described in "Explanation of Attrition Tables B-10 through B-24." The main effect for attrition status was used to answer the question of whether the dropouts differed from the final sample, while the interaction between program status and attrition status was used to test for differential attrition.

Tables B-10 thru B-12 present the results of these analyses for "general attrition." As can be seen, there were only two instances in which a final sample differed significantly from the dropouts on one of the three characteristics:

- Beller's dropouts had a significantly lower mean IQ than his final sample.
- Palmer's dropouts had a significantly lower ISP score (i.e., higher socioeconomic status) than the final sample (although the overall ANOVA was not significant).

In no instances did the final sample and the dropouts differ on more than one of the three characteristics.

There were three instances of differential attrition:

- Beller's control dropouts had a lower mean IQ than any of the other three groups, a bias which would tend to diminish the chances of finding significant differences favoring the program group.
- Gray's control dropouts had the highest ISP score (i.e., lowest SES) while her program dropouts had the lowest ISP (i.e., highest SES).
- Gray's control dropouts had the lowest mother's education.

  These two differences decrease the chances of finding significant differences between program and controls since the lowest SES controls have been lost; however, it is important to note that the control dropouts did not have lower IQ's than the other groups. In fact, they had a slightly higher mean IQ than the control final sample.

Tables B-13 thru B-24 show similar analyses for the individual instruments. As can be seen, there are on each table anywhere from zero to three significants of attrition (either main effects or interaction effects), although these are often from an ANOVA which is not significant overall. In evaluating the evidence of attrition, one must remember the definition of statistical significance - i.e., that the probability of such results occurring by chance is one in twenty. If one regards the significance levels for the main effect of attrition and for the interaction between attrition and program status as two independent tests, then each table (from B-10 to B-24) contains between twelve and twenty independent tests of hypotheses. Looking at the information in this way one can estimate the number of "significant differences" one would have in each table if all occurrences were chance (e.g., in a table with twenty tests of hypotheses, one would expect to find one "significant result" purely by chance). Table B-25 presents this expected number of "significant results" if there were no genuine attrition effect, plus the actual number of significant results. As can be seen, the actual results exceed these expected on the basis of chance in seven of the fifteen tables and are less than those expected by chance in only one table. However, the fact that there are only seven more significant findings than could be expected by chance indicates that the overall effects of attrition are slight.

Two words of caution must be extended in interpreting the attrition results. First, in some instances there were considerable numbers of missing values on the criterion variable (i.e., IQ, ISP or mother's education). Therefore, the analyses may not always represent an accurate assessment of the effects of attrition. Second, the results of attrition analyses can vary depending on how one defines the original sample. These definitions can vary due to the fact that some groups of children were not designated to be followed up at this time, hence, it depends upon one's purposes in doing the analysis whether these groups should be included in the original sample. Definitions of original samples are contained in Table B-5.

The effect of varying definitions of the original sample upon the outcomes of the attrition analyses can be gauged by comparing the results of the attrition presented on the School Record Forms in this appendix with the attrition analysis performed specifically for the discussion of assignment to special education and grade failure. As explained in the section on special education, certain groups from some of the projects were excluded from the analysis _a priori_* and, hence, were not included as part of the original sample in that attrition analysis. These groups were, however, included as part of the original sample in the attrition analysis in the appendix.

Table B-26 presents the results of an attrition analysis performed for the sample as a whole.** This analysis is meant to be exploratory only. Three way ANOVA was performed using the following three factors:

---

* Budget restrictions required that no efforts were made to find some whole cohorts of subject. See also following footnote.

** If no information was received from a project on an instrument, that project was not designated as part of the original sample for that instrument. Also, Karnes' groups 10-14 and Deutsch's non-self-selected controls and Deutsch's waves 5-7 were excluded from the original sample since they were not selected for follow-up at this time.

- which project the child was from
- attrition status
- program status

Again the analysis was performed in hierarchical fashion with project entered first, attrition status second, and program status third. Project was entered first to parcel out the variance attributable to initial differences between projects. Attrition status was entered second because it is the variable of interest and, hence, it is more rigorous (i.e., more likely to arrive at a significant attrition effect) to calculate the main effect of attrition prior to the main effect of program status.

As can be seen from Table B-26, the final sample differs from the dropouts on pretest IQ for all definitions of attrition except attrition on the WISC-R. The only other significant main effect of attrition is on the ISP for the School Record. There were only two significant two-way interactions involving attrition (both between attrition and project): mother's education for the Youth Interview, and ISP for the School record. There were no two-way interactions between attrition and program status (i.e., no evidence for differential attrition as defined in question 3); however, there were six three-way interactions. Overall, it appears in regard to whether a Binet was administered that the total final sample is significantly different from the total dropouts on pretest Binet IQ but there is little evidence for differential attrition between program and control groups.

Attrition in the Original Data

The parameters for the analysis of attrition for each project were defined as follows:

Project	Criterion of Dropout Status	Dependent Variables
Gordon	Existence of a Binet at age 6	Binet at age 3, SES, mother's educatio
ETP	" " " " " " 10	" " pretest " " "
FOHV	" " " " " " 5	" " " 3 " " "
Karnes long	" " " " " " 9	" " pretest " " "
Karnes p-p	" " " " " " 5	" " " " " "
Levenstein	" " " " " " 5	" " age 3 " " "
Miller	" " " " " " 8	" " pretest " " "
Palmer	" " " " " " 5	" " age 3 " " "
Perry	" " " " " " 10	" " pretest " " "
CD	" " " " " " 8	" " " " " "
Carnegie	" " " " " " 7	" " age 3 " " "

Tables B-26 thru B-28 represent the 2 X 2 ANOVA analyses for the differential attrition on the original data. For four projects the significance levels come from different level ANOVA:

- Karnes pre-post: 2 X 5 groups
- Karnes longitudinal: 2 X 5 groups
- Miller program comparison: 2 X 4 groups
- Curriculum Demonstration: 2 X 3 groups

As can be seen, there are very few indications of differential attrition. Only the Levenstein project on the ISP has a significant attrition effect which contributes to an overall significant ANOVA. Other projects did show some significant attrition effects; however, their overall ANOVA's were not significant.

One should note that the attrition reported here for Levenstien is artificial due to the stipulation of this analysis that a subject must have a Binet score at the designated posttest in order to be considered part of the final sample. Much of Levenstein's "attrition" is caused by the fact that many of her subjects were given WISC's rather than Binets in later follow-up periods.

Conclusions

The overall evidence seems to indicate that, although there are some instances of differential attrition, the final samples are representative of the original samples. Overall (general attrition) there were no significant differences between the demographic characteristics of those subjects found and those not found. Such attrition as was significant dealt with whether specific pieces of data were in hand on July 1, 1977. Thus, attrition appears to have been a fairly random process. This conclusion is not surprising given the fact that much of the attrition occurred while neither program nor control children were receiving any major benefits (or costs) from the project (i.e., after project termination). The major reason to suspect biases caused by attrition is that the benefits received from the program will cause a different pattern of attrition in the program versus control groups. Since the current follow-up occurred several years after program termination, there is little reason to susp-ct that differential patterns of attrition would occur when neither group was receiving benefits from the program. The most crucial period for which attrition must be investigated in evaluation research is the period of program duration. The question of attrition during the program has been addressed by most of the projects in their original reports which deal with more immediate posttests. The fact that less than 3% of the parents and/or youths contacted refused to participate in the follow-up eliminates any concerns we had about contemporaneous differential attrition.

Explanation of Attrition Tables B-10 to B-24, B-27, and B-28

Separate tables are provided for attrition on each of four instruments in the follow-up (Parent Interview, Youth Interview, WISC, and School Record) and for those who had been given any instrument versus those not found at all (the traditional concept of attrition). Analyses were performed for each project on three dependent variables: Stanford-Binet or PPVT pretest IQ score, Hollingshead Index of Social Position score, and mother's education. Thus there are a total of fifteen tables (three dependent variables time five measures of attrition status).

All tables are based on separate two times two ANOVA performed for each project.* The first factor in the ANOVA is attrition status (dropout versus final sample), while the second is program status (program vs. control). ANOVA is performed using a hierarchical method for assigning variance (SPSS Option Ten) with attrition status the first variable entered (i.e. the F statistic for the attrition main effect is computed without controlling for program status). The first three columns in the tables present three significance levels taken from th 2 X 2 ANOVA's. The first is the significance level for the main effet of attrition status; the second, the significance level for the interaction between attrition status and program status; the third, the overall significance level for the 2 X 2 ANOVA. They are taken fom a standard SPSS ANOVA printout. For example, the numbers for the Gra project in Table B-5 were taken from the table on the following page.

---

* In two projects which were designed to compare different curricula, the attrition analyses are performed by 2 X 5 (Karnes) and 2 X 4 (Miller program only) ANOVA's. These analyze differential attrition among different program groups. Only the significance levels are reported.

The last four columns give the mean score on the dependent variables for each of the four cells defined by the 2 X 2 ANOVA, with sample size for each cell given in parentheses below the mean score.

The tables do not present all of the information given in a standard ANOVA table simply because the entire analysis consists of almost 150 separate ANOVA's (three dependent variables times five measures of attrition times ten projects).

The last row in each table gives the ANOVA results when all projects are grouped together.

The tables are organized as follows:

B-5 pretest IQ by Parent Interview status for ten projects

B-6 ISP by Parent Interview status for ten projects

B-7 mother's education by Parent Interview status for ten projets

B-8 pretest IQ by Parent Interview status for ten projects

B-9 ISP by Parent Interview status for ten projects etc.

Example: SPSS ANOVA Printout for GRAY

```
FILE    KIDS       (CREATION DATE = 07/09/77)
SUBFILE    GETP

* * * * * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * * * * *
                       SBIQPRE
                    BY APAIN
                       TC
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *


                                             SUM OF                      MEAN            SIGNIF
SOURCE OF VARIATION                          SQUARES        DF          SQUARE        F     OF F

MAIN EFFECTS                                  96.320         2          48.160    0.267    0.999
   APAIN                                       0.482         1           0.482    0.003    0.999   Col. 1
   TC                                         95.837         1          95.837    0.532    0.999

2-WAY INTERACTIONS                             2.285         1           2.285    0.013    0.999
   APAIN    TC                                 2.285         1           2.285    0.013    0.999   Col. 2

EXPLAINED                                     98.605         3          32.868    0.182    0.999
                                                                                                   Col. 3
RESIDUAL                                   15136.145        84         180.192

TOTAL                                      15234.750        87         175.112


      92 CASES WERE PROCESSED.
       4 CASES (  4.3 PCT) WERE MISSING.
```

Attrition in the Special Education and Grade Failure Analysis

Due to the importance of the findings on special education and grade failure, additional attrition analyses were conducted specifically for the groups used in these analyses. Separate attrition analyses are presented for each project.

Ira Gordon's Parent/Education/Backyard Learning Center Project: The attrition analysis for the Gordon project is reported in "School Performance as a Function of Early Stimulation."(13) Bayley Mental Development Index scores at age two and Stanford-Binet IQ scores at ages three and six were compared between the 1976 sample and the sample at the end of intervention for each of the groups. No significant differences were found.

Susan Gray's Early Training Project: The final sample and the dropouts were compared on three measures: Binet pretest IQ score, head of household socioeconomic status (Hollingshead Index), and mother's education. T-tests were run for the full sample and for program and control groups separately. No significant differences were found. Two by two analyses of variance drop-out status x program status) for each variable also yielded no significant differences.

Both t-tests and ANOVA were run because of the problems in assigning main effects in ANOVA of unequal cell sizes. The t-tests assign to dropout status all variance which could be attributed to dropout status, while the ANOVA tests for interaction between dropout status and program status.

Merle Karnes' Research and Development Program on Preschool Children: For each of three measures (pretest Stanford-Binet IQ, mother's education, and Hollingshead SES) a 5 X 2 ANOVA (with the five experimental groups as one factor and dropout status as the other) was performed. No significant differences were found. The t-tests comparing the final sample with the dropouts on the above measures for each of the experimental groups individually also revealed no significant difference.

Phyllis Levenstein's Mother-Child Home Program: The Levenstein project, which has followed the children almost annually since the inception of the program had data on all but one of the program children and of all of the control children. Thus, there were no significant differences on any characteristic since there was no variance in the dropout category.

Louise Miller's Experimental Variation of Head Start Curricula: The final sample and the dropouts were compared on pretest Stanford-Binet IQ and on mother's education using identical procedures to those used in the Gray project [Socioeconomic status (SES) scores were not available for the control group]. No significant differences were found.

A second analysis (4 X 2 ANOVA with the four experimental groups as one factor and dropout status as the other) revealed no significant differences among experimental groups with respect to pretest IQ, mother's education, or Hollingshead SES.

Francis Palmer's Harlem Training Project: The attrition analysis reported by Palmer concludes:

> Attrition analysis, conducted by comparing IQ's and social class at ages three and five (the last assessment) for those found and not found in 1975, showed no significant differences between any of the groups in the design. (14)

David Weikart's Perry Preschool Project: The Weikart project was able to find school record data on all 123 children.

Edward Zigler's New Haven Follow-Through Study: Two measures (PPVT IQ scores at age five and mother's education) were compared between the dropouts and the final sample by t-tests for the total sample and for the program and control groups separately. Two by two ANOVA's (dropout status and program status) were also computed. No significant differences were found on the t-tests. However, the ANOVA on PPVT IQ showed a significant interaction between dropout status and experimental-control status (overall F=3.128 p=.028). The control dropouts had the lowest average IQ, while the program dropouts had the highest average IQ. Hollingshead SES measures were not available.

Table B-5: Final Sample as a Percent of Original Sample by Project: "General Attrition"

Project	Final Controls as % of Original Controls	Final Program as % of Original Program	Total Final Sample as % of Original Total Sample	χ^2	Significance	Definition of Original Sample
Beller	48.6* (54)	54.2* (32)	50.6 (86)	0.284	.5943	full 170 cases sent to Cornell
Deutsch	6.3 (12)	16.3 (51)	12.5 (63)	10.173	.0014	first 4 waves, groups 1 & 2 only = 504 cases
Gordon	33.8 (23)	30.3 (73)	31.1 (96)	0.166	.6835	all 309 cases with test scores sent to Cornell
Gray	85.4 (41)	81.8 (36)	83.7 (77)	0.03394	.8538	all 92 cases ETP cases sent to Cornell
Karnes		82.9 (121)	82.9 (121)	-----	-----	groups 1 thru 10 = 146 cases
Levenstein	75.6 (59)	73.8 (127)	74.4 (186)	0.021	.8836	all 250 cases sent to Cornell (first 5 waves
Miller	56.1 (32)	50.9 (109)	52.0 (141)	.0011	.9739	all 270 cases sent to Cornell
Palmer	70.1 (47)	71.9 (174)	71.5 (221)	0.016	.8980	all 309 cases sent to Cornell
Weikart	100.0 (65)	100.0 (58)	100.0 (123)	-----	-----	all 123 Perry cases sent to Cornell

---

* Since these figures were calculated, Beller has located additional subjects, bringing those figures up to 75% and 87% respectively. These additional data did not reach us in time for this analysis, and will be included in future reports. However, note that there are no significant differences between the sample which was on hand and the original groups. The delay in the location of the Beller subjects was due to an unusually complex decision process in the public schools.

Table B-6: Final Sample as a Percent of Original Sample of Parent Interview, by Project

Project	Final Controls as a % of Original Controls	Final Program as a % of Original Program	Final Total as a % of Original Total	X^2	Significance
Beller	42.3 (47)	52.5 (31)	45.9 (78)	1.2295	.2675
Deutsch	3.1 (6)	13.1 (41)	9.3 (47)	12.9418	.0003
Gordon	33.8 (23)	30.3 (73)	31.1 (96)	0.1662	.6835
Gray	75.0 (36)	81.8 (36)	78.3 (72)	0.2905	.5899
Karnes		82.2 (120)	82.2 (120)	-----	-----
Levenstein	21.8 (17)	39.0 (67)	33.6 (84)	6.3335	.0118
Miller	56.1 (32)	50.9 (109)	52.0 (141)	.00107	.9739
Palmer	41.8 (28)	40.9 (99)	41.1 (127)	0.0001	.9917
Weikart	86.2 (56)	86.2 (50)	86.2 (106)	0.0641	.8001

Table B-7: Final Sample as a Percent of Original for Youth Interview by Project

Project	Final Controls as a % of Original Controls	Final Program as a % of Original Program	Final Total as a % of Original Total	x^2	Significance
Beller	33.3 (37)	37.3 (22)	34.7 (59)	0.1200	.7290
Deutsch	5.7 (11)	14.7 (46)	11.3 (57)	8.7512	.0031
Gordon	19.1 (13)	12.0 (29)	13.6 (42)	1.7033	.1919
Gray	75.0 (36)	75.0 (33)	75.0 (69)	0.0581	.8096
Karnes		78.1 (114)	78.1 (114)	-----	-----
Levenstein	23.1 (18,	23.8 (41)	23.6 (59)	0.0009	.9764
Miller	56.1 (32)	50.9 (109)	52.0 (141)	0.0011	.9739
Palmer	43.3 (29)	40.9 (99)	41.4 (128)	0.0437	.8344
Weikart	87.7 (57)	81.0 (47)	84.6 (104)	0.5929	.4413

Table B-8: Final Sample as a Percent of Original for School Record Form by Project

Project	Final Controls as a % of Original Controls	Final Program as a % of Original Program	Final Total as a % of Original Total	X^2	Significance
Beller	19.8 (22)	32.2 (19)	24.1 (41)	2.5869	.1078
Deutsch	2.1 (4)	9.3 (29)	6.5 (33)	8.9578	.0028
Gordon	0 (0)	0 (0)	0 (0)	-----	-----
Gray	79.2 (38)	81.8 (36)	80.4 (74)	0.0033	.9544
Karnes		72.6 (107)	72.6 (107)	-----	-----
Levenstein	71.8 (56)	37.2 (64)	48.0 (120)	24.3505	.0000
Miller	56.1 (32)	50.0 (107)	51.3 (139)	0.0180	.8934
Palmer	67.2 (45)	63.2 (153)	64.1 (198)	0.2036	.6519
Weikart	100.0 (65)	100.0 (58)	100.0 (123)	-----	-----

Table B-9: Final Sample as a Percent of Original for WISC-R by Project

Project	Final Controls as a % of Original Controls	Final Program as a % of Original Program	Final Total as a % of Original Total	X^2	Significance
Beller	0 (0)	0 (0)	0 (0)	-----	-----
Deutsch	0 (0)	0 (0)	0 (0)	-----	-----
Gordon	0 (0)	0 (0)	0 (0)	-----	-----
Gray	79.2 (38)	77.3 (34)	78.3 (72)	0.0011	.9737
Karnes		76.7 (112)	76.7 (112)	-----	-----
Levenstein	32.1 (25)	29.7 (51)	30.4 (76)	0.1074	.7432
Miller	56.1 (32)	50.9 (109)	52.0 (141)	0.0011	.9739
Palmer	37.3 (25)	38.8 (94)	38.5 (119)	0.0074	.9316
Weikart	86.2 (56)	93.1 (54)	89.4 (110)	0.9172	.3382

Table B-10: General Attrition - Comparison of Pretest Stanford-Binet IQ Scores

Project	Significance Levels			Mean I.Q. Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout		Final Sample	
				Program	Control	Program	Control
Beller	.003	.030	.001	92.32 (25)	84.11 (57)	92.03 (32)	92.91 (54)
Deutsch	.999	.193	.999	92.05 (229)	91.68 (120)	93.51 (45)	87.00 (9)
Gordon *	.999	.999	.296	95.38 (80)	92.90 (31)	94.27 (64)	89.19 (21)
Gray	.999	.999	.999	89.88 (8)	88.71 (7)	89.33 (36)	87.08 (37)
Karnes	.999	.999	.999				
Levenstein **	.999	.177	.999	81.79 (42)	86.67 (9)	84.50 (125)	82.89 (27)
Miller Program/Con.	.306	.999	.308	92.09 (104)	87.81 (16)	93.62 (109)	90.44 (18)
Miller Program/Comp.	.999	.999	.220				
Palmer	.111	.056	.001	90.62 (34)	86.89 (19)	95.92 (86)	84.64 (44)
Weikart	No Dropouts on School Record					79.57 (58)	78.54 (65)

* Stanford-Binet at age three (a posttest IQ) was used in this analysis

** PPVT pretest I.Q.

*** Only half the program group was given a pretest

Table B-11: General Attrition – Comparison of Hollingshead ISP Scores*

Project	Significance Levels			Mean ISP Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout		Final Sample	
				Program	Control	Program	Control
Beller	No ISP data						
Deutsch	.310	.999	.999	66.71 (45)	66.05 (74)	65.05 (40)	62.67 (3)
Gordon	No ISP data	received on	controls				
Gray	.999	.007	.058	63.60 (5)	74.14 (7)	70.52 (31)	69.00 (39)
Karnes	.999	.999	.999				
Levenstein	.999	.097	.315	66.23 (44)	63.79 (19)	64.17 (125)	66.00 (59)
Miller (Program & Control)	No ISP data	on controls					
Miller (Program Comp.)	.999	.999	.999				
Palmer	.024	.999	.105	56.75 (64)	52.85 (20)	59.12 (164)	59.04 (47)
Weikart	No Dropouts	on School Record				67.81 (58)	69.09 (64)

*The Hollingshead Index of Social Position (ISP) ranges from 11 to 77, with "11" representing the highest social class and "77" representing the lowest social class.

Table B-12: General Attrition - Comparison of Mother's Education

	Significance Levels			Mean Grades Completed			
				Dropout		Final Sample	
Project	Attrition Main Effect	Interaction	Overall ANOVA	Program	Control	Program	Control
Beller	No education data						
Deutsch	.999	.097	.999	10.36 (45)	10.03 (74)	9.95 (41)	12.00 (3)
Gordon	Education data received on only two controls						
Gray	.157	.018	.048	9.67 (6)	7.00 (7)	8.79 (34)	9.68 (41)
Karnes	.999	.335	.411				
Levenstein	.999	.139	.999	10.27 (44)	10.84 (19)	10.65 (126)	10.27 (59)
Miller Program/Con.	.999	.999	.999	10.89 (95)	10.67 (15)	10.87 (100)	10.29 (17)
Miller Program/Comp.	.999	.999	.999				
Palmer	.158	.999	.999	11.45 (65)	11.50 (20)	11.12 (167)	11.23 (47)
Weikart	No Dropouts on School Record					9.47 (58)	9.38 (65)

Table B-13: Attrition on the Parent Interview - Comparison of Pretest Stanford-Binet IQ Scores

	Significance Levels			Mean I.Q. Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout		Final Sample	
Project				Program	Control	Program	Control
Beller	.079	.186	.067	92.46 (26)	86.22 (64)	91.90 (31)	91.34 (47)
Deutsch	.999	.194	.999	92.08 (239)	91.59 (124)	93.74 (35)	85.40 (5)
Gordon *	.999	.999	.296	95.38 (80)	92.90 (31)	94.27 (64)	89.19 (21)
Gray	.999	.999	.999	89.88 (8)	87.17 (12)	89.33 (36)	87.41 (32)
Karnes	.999	.999	.999				
Levenstein **	.026	-----	.062	82.13 (100)	83.82 (36)	86.33 (67)	(0)
Miller Program/Con.	.306	.999	.308	92.09 (104)	87.82 (16)	93.62 (109)	90.44 (18)
Miller Program/Comp.	.999	.999	.220				
Palmer ***	.999	.999	.001	93.21 (72)	85.42 (36)	96.23 (48)	85.19 (27)
Weikart	.999	.999	.999	81.00 (8)	78.00 (9)	79.34 (50)	78.63 (56)

* Stanford-Binet at age three (a posttest IQ) was used in this analysis.

** PPVT pretest I.Q.

*** Only half of the program group was given a pretest.

Table B-14: Attrition on the Parent Interview - Comparison of Hollingshead ISP Scores

	Significance Levels			Mean ISP Scores			
	Attrition		Overall	Dropout		Final Sample	
Project	Main Effect	Interaction	ANOVA	Program	Control	Program	Control
Beller	No ISP data						
Deutsch	.152	.999	.999	67.02 (52)	65.93 (76)	64.21 (33)	65.00 (1)
Gordon	No ISP data	received on	controls				
Gray	.999	.002	.019	63.60 (5)	74.09 (11)	70.52 (31)	68.43 (35)
Karnes	.999	.999	.999				
Levenstein	.156	.213	.222	64.33 (102)	64.56 (61)	65.27 (67)	68.71 (17)
Miller Program/Con.	.306	.999	.308				
Miller Program/Comp.	.999	.999	.999				
Palmer	.999	.999	.999	58.10 (133)	56.69 (39)	58.95 (95)	57.89 (28)
Weikart	.306	.999	.999	65.13 (8)	68.00 (9)	68.24 (50)	69.27 (55)

* The Hollingshead ISP (Index of Social Position) ranges from 11 to 77 with "11" representing the highest social class and "77" representing the lowest social class.

Table B-15: Attrition on the Parent Interview – Comparison of Mother's Education

Project	Significance Levels			Mean Grades Completed			
				Dropout		Final Sample	
	Attrition Main Effect	Interaction	Overall ANOVA	Program	Control	Program	Control
Beller	No Education data						
Deutsch	.999	.269	.999	10.43 (53)	10.08 (76)	9.73 (33)	12.00 (1)
Gordon	Education data received on only two controls						
Gray	.062	.014	.018	9.67 (6)	7.42 (12)	8.79 (34)	9.92 (36)
Karnes	.999	.340	.416				
Levenstein	.076	.999	.276	10.74 (103)	10.56 (61)	10.27 (67)	9.88 (17)
Miller Program/Con.	.999	.999	.999	10.67 (95)	10.89 (15)	10.87 (100)	10.29 (17)
Miller Program/Comp.	.999	.999	.999				
Palmer	.999	.999	.999	11.30 (134)	11.26 (39)	11.09 (98)	11.39 (28)
Weikart	.999	.058	.211	8.00 (8)	9.78 (9)	9.70 (50)	9.32 (56)

Table B-16: Attrition on the Youth Interview - Comparison of Pretest Stanford-Binet IQ Scores

Project	Significance Levels			Mean I.Q. Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout		Final Sample	
				Program	Control	Program	Control
Beller	.011	.020	.003	92.71 (35)	85.49 (74)	91.27 (22)	94.19 (37)
Deutsch	.999	.999	.999	91.94 (233)	91.48 (121)	94.29 (41)	89.38 (8)
Gordon*	.999	.999	.367	95.27 (116)	90.98 (40)	93.29 (28)	92.83 (12)
Gray	.999	.999	.999	88.73 (11)	88.09 (11)	89.67 (33)	87.09 (33)
Karnes	.265	.999	.999				
Levenstein**	.068	---	.175	82.88 (126)	83.83 (36)	86.68 (41)	(0)
Miller Program/Con.	.306	.999	.308	92.09 (104)	87.81 (16)	93.62 (109)	90.44 (18)
Miller Program/Comp.	.999	.999	.220				
Palmer***	.999	.999	.001	93.29 (73)	85.17 (35)	96.17 (47)	85.50 (28)
Weikart	.999	.999	.999	80.55 (11)	79.75 (8)	79.34 (47)	78.36 (57)
							181

* Stanford-Binet at age three (a posttest IQ) was used in this analysis.

** PPVT pretest I.Q.

*** Only half the program group was given a pretest.

Table 17: Attrition on the Youth Interview - Comparison of Hollingshead ISP Scores*

	Significance Levels			Mean ISP Scores			
				Dropout		Final Sample	
Project	Attrition Main Effect	Interaction	Overall ANOVA	Program	Control	Program	Control
Beller	No ISP data						
Deutsch	.999	.999	.999	66.04 (48)	66.05 (74)	65.78 (37)	62.67 ([illegible])
Gordon	No ISP data received on Controls						
Gray	.999	.227	.999	68.13 (8)	71.75 (12)	69.96 (28)	69.09 (34)
Karnes	.999	.999	.999				
Levenstein	.271	.080	.196	64.72 (128)	64.42 (60)	64.66 (41)	68.94 (18)
Miller Program/Con.	No ISP data on Controls						
Miller Program/Comp.	.999	.999	.999				
Palmer	.999	.999	.999	58.26 (133)	56.66 (38)	58.72 (95)	57.90 (29)
Weikart	.016	.999	.081	63.09 (11)	66.50 (8)	68.91 (47)	69.46 (56)

---

* The Hollingshead Index of Social Position (ISP) ranges from 11 to 77, with "11" representing the highest social class and "77" representing the lowest social class.

Table B-18: Attrition on the Youth Interview - Comparison of Mother's Education

	Significance Levels			Mean Grade Completed			
				Dropout		Final Sample	
Project	Attrition Main Effect	Interaction	Overall ANOVA	Program	Control	Program	Control
Beller	No education data						
Deutsch	.999	.269	.999	10.42 (48)	10.03 (74)	9.84 (38)	12.00 (3)
Gordon	Education data received on only two controls						
Gray	.001	.172	.003	8.00 (9)	7.17 (12)	9.19 (31)	10.00 (36)
Karnes	.999	.999	.999				
Levenstein	.012	.999	.088	10.73 (129)	10.60 (60)	10.00 (41)	9.78 (18)
Miller Program/Con.	.999	.999	.999	10.89 (95)	10.67 (15)	10.87 (100)	10.29 (17)
Miller Program/Coup.	.999	.999	.999				
Palmer	.999	.999	.999	11.31 (134)	11.29 (38)	11.07 (98)	11.34 (29)
Weikart	.303	.037	.142	8.18 (11)	10.00 (8)	9.77 (47)	9.30 (57)

Table B-19: Attrition on the School Record - Comparison of Pretest Stanford-Binet IQ Scores

Project	Significance Levels			Mean I.Q. Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout		Final Sample	
				Program	Control	Program	Control
Beller	.002	.005	.001	92.39 (38)	85.92 (89)	91.68 (19)	98.36 (22)
Deutsch	.117	.999	.324	91.88 (248)	91.40 (125)	96.23 (26)	89.75 (4)
Gordon *	No school records received						
Gray	.999	.999	.999	89.88 (8)	86.40 (10)	89.33 (36)	87.62 (34)
Karnes	.999	.313	.999				
Levenstein **	.999	.999	.999	84.26 (105)	86.66 (9)	83.06 (62)	82.89 (27)
Miller Program/Con.	.999	.999	.319	92.14 (106)	87.81 (16)	93.60 (107)	90.44 (18)
Miller Program/Comp.	.999	.999	.228				
Palmer	.105	.194	.001	91.10 (42)	85.48 (21)	96.21 (78)	85.24 (42)
Weikart	No Dropouts					79.57 (58)	78.54 (65)

* Stanford-Binet at age three (a posttest IQ) was used in this analysis

** PPVT pretest I.Q.

*** Only half the program group was given a pretest

Table B-20: Attrition on the School Record - Comparison of the Hollingshead ISP Scores*

	Significance Levels			Mean ISP Scores			
	Attrition		Overall	Dropout		Final Sample	
Project	Main Effect	Interaction	ANOVA	Program	Control	Program	Control
Beller	No ISP data						
Deutsch	.999	---	.999	66.36 (61)	65.92 (77)	64.83 (24)	(0)
Gordon	No ISP data	received on	controls				
Gray	.999	.030	.175	63.60 (5)	71.50 (10)	70.52 (31)	69.31 (36)
Karnes	.999	.999	.999				
Levenstein	.999	.999	.999	65.18 (107)	65.05 (22)	63.89 (62)	65.63 (56)
Miller Program/Con.	No ISP data	on controls					
Miller Program/Comp.	.999	.999	.999				
Palmer	.006	.999	.043	55.76 (79)	53.82 (22)	58.84 (45)	59.88 (149)
Weikart	No dropouts					67.81 (58)	69.09 (64)

---

* The Hollingshead Index of Social Position (ISP) ranges from 11 to 77 , with "11" representing the highest social class and "77" representing the lowest social class.

Table B-21: Attrition on the School Record - Comparison of Mother's Education

Project	Significance Levels			Mean Grades Completed			
				Dropout		Final Sample	
	Attrition Main Effect	Interaction	Overall ANOVA	Program	Control	Program	Control
Beller	No Education data						
Deutsch	.143	-----	.249	10.42 (61)	10.10 (77)	9.52 (25)	(0)
Gordon	Education data received on only two controls						
Gray	.999	.065	.182	9.67 (6)	7.90 (10)	8.79 (34)	9.66 (38)
Karnes	.045	.999	.261				
Levenstein	.999	.999	.999	10.43 (107)	10.23 (22)	10.76 (63)	10.48 (56)
Miller Program/Con.	.999	.999	.999	10.88 (97)	10.67 (15)	10.89 (98)	10.29 (17)
Miller Program/Comp.	.999	.999	.999				
Palmer	.198	.999	.999	11.42 (81)	11.50 (22)	11.10 (151)	11.22 (45)
Weikart	No Dropouts					9.47 (58)	9.38 (65)

Table B-22: Attrition on the WISC-R – Comparison of Pretest Stanford-Binet IQ Scores

Project	Significance Levels			Mean I.Q. Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout Program	Dropout Control	Final Sample Program	Final Sample Control
Beller	No WISC-R's received						
Deutsch	No WISC-R's received						
Gorden *	No WISC-R's received						
Gray	.999	.999	.999	91.40 (10)	86.40 (10)	88.85 (34)	87.62 (34)
Karnes	.217	.999	.999				
Levenstein**	.036	-----	.096	82.74 (117)	83.83 (36)	86.32 (50)	(0)
Miller Program/Con.	.306	.999	.308	92.09 (104)	87.81 (16)	93.62 (109)	90.44 (18)
Miller Program/Comp.	.999	.999	.220				
Palmer	.216	.999	.001	93.25 (75)	84.87 (39)	96.36 (45)	86.04 (24)
Weikart	.174	.999	.306	80.00 (4)	81.89 (9)	79.54 (54)	78.00 (56)

* Stanford-Binet at age three (a posttest IQ) was used in this analysis

** PPVT pretest I.Q.

*** Only half the program group was given a pretest

Table B-23: Attrition on the WISC-R - Comparison of the Hollingshead ISP Scores

	Significance Levels			Mean ISP Scores			
	Attrition		Overall	Dropout		Final Sample	
Project	Main Effect	Interaction	ANOVA	Program	Control	Program	Control
Beller	No ISP data						
Deutsch	No WISC's research						
Gordon	No ISP data received on controls						
Gray	.999	.156	.999	66.86 (7)	71.50 (10)	70.21 (29)	69.31 (36)
Karnes	.999	.999	.999				
Levenstein	.999	.032	.174	65.21 (119)	64.36 (53)	63.50 (50)	67.80 (25)
Miller Program/Con.	No ISP data on controls						
Miller Program/Comp.	.999	.999	.999				
Palmer	.006	.999	.043	57.50 (138)	56.33 (42)	59.91 (90)	58.64 (25)
Weikart	.999	.999	.999	67.50 (4)	67.11 (9)	67.83 (54)	69.42 (55)

* The Hollingshead Index of Social Position (ISP) ranges from 11 to 77, with "11" representing the highest social class and "77" representing the lowest social class.

Table B-24: Attrition on the WISC-R - Comparison of Mother's Education

Project	Significance Levels			Mean Years of Education			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout		Final Sample	
				Program	Control	Program	Control
Beller	No Education data						
Deutsch	No WISC's received						
Gordon	No WISC's received						
Gray	.110	.253	.232	8.75 (8)	7.90 (10)	8.97 (32)	9.66 (38)
Karnes	.999	.219	.329				
Levenstein	.999	.999	.999	10.62 (120)	10.55 (53)	10.40 (50)	10.12 (25)
Miller Program/Con.	.999	.999	.999	10.89 (95)	10.67 (15)	10.87 (100)	10.29 (17)
Miller Program Comp.	.999	.999	.999				
Palmer	.999	.999	.999	11.29 (139)	11.26 (42)	11.10 (93)	11.40 (25)
Weikart	.161	.999	.999	10.50 (4)	10.11 (9)	9.39 (54)	9.27 (56)

Table B-25: Expected Number of Chance "Significant Findings" and Actual Number of Significant Findings for Tables B-10 to B-24

Table	Expected Number of "Significant Findings"*	Actual Number of Significant Findings**	Difference
B-10	1	2	+1
B-11	1	2	+1
B-12	1	1	
B-13	1	1	
B-14	1	1	
B-15	1	1	
B-16	1	2	+1
B-17	1	1	
B-18	1	3	+2
B-19	1	2	+1
B-20	1	2	+1
B-21	1	1	
B-22	1	1	
B-23	1	2	+1
B-24	1	0	-1

* If there were no attrition, i.e., by chance. Rounded to nearest integer.

** On Attrition Main Effect or Interaction Attrition X Program Status.

Table B-26: Results of 3 Way ANOVA: Significance Levels

	Effects Involving Attrition				Other Main Effects		Interaction Project X Program Status	Overall ANOVA	D.F.
	Attrition Main Effect	Interactions			Project Main Effect	Program Status			
		Project X Attrition	Program Status X Attrition	3 Way					
Parent Interview									
pretest I.Q.	.034	.999	.999	.999	.001	.001	.024	.001	1402
ISP	.252	.999	.999	.185	.001	.999	.999	.001	1320
Mother's Ed.	.999	.119	.999	.004	.001	.999	.999	.001	1401
Youth Interview									
pretest I.Q.	.011	.999	.999	.187	.001	.001	.017	.001	1402
ISP	.280	.999	.999	.999	.001	.999	.999	.001	1320
Mother's Ed.	.999	.001	.999	.031	.001	.999	.999	.001	1401
School Record									
pretest I.Q.	.001	.999	.999	.036	.001	.001	.020	.001	1402
ISP	.041	.052	.999	.217	.001	.999	.999	.001	1212
Mother's Ed.	.305	.140	.999	.151	.001	.999	.999	.001	1286
WISC-R									
pretest I.Q.	.106	.999	.999	.999	.001	.001	.038	.001	831
ISP	.130	.999	.999	.313	.001	.999	.999	.001	1031
Mother's Ed.	.999	.196	.999	.999	.001	.999	.999	.001	1102
"General"									
pretest I.Q.	.003	.416	.999	.043	.001	.001	.045	.001	1402
ISP	.175	.190	.999	.090	.001	.999	.999	.001	1320
Mother's Ed.	.999	.999	.999	.003	.001	.999	.999	.001	1401

Table B-27: Attrition on the Original Data - Comparison of Stanford-Binet Pretest IQ Scores

	Significance Levels			Mean I.Q. Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout		Final Sample	
				Program	Control	Program	Control
Beller	.999	.999	.272	96.57 (7)	88.58 (12)	91.54 (50)	88.36 (99)
Deutsch	.019	.999	.094	91.94 (256)	91.07 (126)	97.28 (18)	103.00 (3)
Gordon*	.999	.999	.999	94.17 (30)	92.00 (7)	95.07 (114)	91.31 (45)
Gray ETP	.246	.035	.100	78.00 (6)	90.67 (6)	91.24 (38)	86.82 (38)
Gray FOHV*	.058	.999	.130	75.50 (2)	63.00 (1)	88.19 (27)	82.54 (13)
Karnes P-P	.999	-----	.001				
Karnes Long	.999	.999	.999				
Levenstein **	.263	.117	.299	82.59 (97)	85.00 (26)	85.51 (70)	80.80 (10)
Miller P-C	.999	.999	.999	93.38 (40)	90.00 (5)	92.76 (173)	89.07 (29)
Miller P only	.999	.999	.229				
Palmer ***	.999	.110	.001	91.80 (30)	89.41 (17)	93.29 (198)	83.80 (46)
Perry	.243	.999	.999	81.00 (1)	81.50 (8)	79.54 (57)	78.12 (57)
C. Dem	.159	-----	.999				
Carnegie	.250	-----	.282	91.50 (2)	----- (0)	105.50 (18)	100.44 (18)

* Stanford-Binet at age three (a posttest IQ) was used in this analysis

** PPVT pretest I.Q.

*** Only half the program group was given a pretest.

Table B-28: Attrition on the Original Data - Comparison of ISP Scores*

	Significance Levels			Mean ISP Scores			
	Attrition Main Effect	Interaction	Overall ANOVA	Dropout Program	Dropout Control	Final Sample Program	Final Sample Control
Beller	No ISP data						
Deutsch	.252	.999	.999	65.22 (63)	65.93 (69)	67.95 (22)	65.88 (8)
Gordon	No ISP data on Controls						
Gray ETP	.320	.999	.999	73.00 (4)	71.00 (6)	69.13 (32)	69.6 (40)
Gray FOHV	.999	.999	.329	61.67 (3)	69.00 (1)	65.52 (27)	68.80 (15)
Karnes P-P	.114	-----	.023				
Karnes Long	.019	.999	.389				
Levenstein	.025	.041	.027	66.66 (73)	65.15 (72)	63.17 (96)	63.22 (6)
Miller P-C	No ISP data on Controls			61.70 (27)		63.42 (143)	
Miller P only	.999	.999	.999				
Palmer	.071	.274	.211	53.81 (27)	56.84 (19)	59.07 (201)	57.33 (48)
Perry	.999	.999	.999	69.00 (1)	67.13 (8)	67.79 (56)	69.38 (57)
C. Dem	.999	.999	.999				
Carnegie	No ISP scores available						

---

* The Hollingshead Index of Social Position (ISP), ranges from 11 to 77, with "11" representing the highest social class and "77" representing the lowest social class.

Table B-29: Attrition on the Original Data - Comparison of Mother's Education

	Significance Levels			Mean Grades Completed			
				Dropout		Final Sample	
	Attrition Main Effect	Interaction	Overall ANOVA	Program	Control	Program	Control
Beller	No Mother's Education Data						
Deutsch	.999	.211	.999	10.34 (64)	10.04 (69)	9.64 (22)	10.63 (8)
Gordon	No Mother's Education Data						
Gray ETP	.190	.999	.999	8.40 (5)	8.00 (6)	9.00 (35)	9.48 (42)
Gray FOHV	.999	.999	.260	10.67 (3)	10.00 (1)	10.53 (30)	9.40 (15)
Karnes P-P	.164	-----	.999				
Karnes Long	.050	.341	.377				
Levenstein	.999	.999	.999	10.60 (73)	10.40 (72)	10.50 (97)	10.52 (6)
Miller P-C	.999	.209	.353	10.86 (36)	11.40 (5)	10.89 (159)	10.30 (27)
Miller P only	.999	.222	.298				
Palmer	.999	.999	.999	11.19 (27)	11.53 (19)	11.21 (205)	11.23 (48)
Perry	.120	.999	.380	10.00 (1)	8.13 (8)	9.46 (57)	9.56 (57)
C. Dem	.999	-----	.143				
Carnegie	.042	.264	.054	11.50 (2)	14.00 (1)	10.30 (20)	10.82 (33)

Appendix B References

Sample Selection and Assignment; Attrition

1. Ilene Bernstein, et al., "External Validity and Evaluative Research; A Codification of Problems" in Ilena Bernstein (ed), *Validity Issues in Evaluative Research*, Sage Contemporary Social Science Issues 23, 1976.

2. Ira Gordon, "A Parent Education Approach to Provision of Early Stimulation for the Culturally Disadvantaged," Institute for Development of Human Resources, 1967, pp. 17-22.

3. *Ibid.*, p. 22.

4. Ira Gordon, "Early Child Stimulation Through Parent Education," Institute for Development of Human Resources, 1969, p. 15.

5. Ira Gordon and Barry Guinaugh, "A Home Learning Center Approach to Early Stimulation," Institute for Development of Human Resources, p. 10.

6. Susan Gray and Rupert Klaus, "The Early Training Project: A Seventh-Year Report," Child Development, Vol. 41, #4, Dec., 1970, p. 910.

7. Merle Karnes, "Research and Development Program on Preschool Disadvantaged Children, Final Report," Contract No. OE 6-10-235, U.S. Office of Education, p. 59.

8. Phyllis Levenstein, "Verbal Interaction Project," Family Service Association of Nassau County, Inc. Mineola, NY, February, 1971, pp. 2-7.

9. Louise Miller and Jean Dyer, *Four Preschool Programs, Their Dimensions and Effects, Monographs of the Society for Research in Child Development*, Vol. 40, 1975, pp. 15-16.

10. Francis Palmer, *Final Report* HD02253 and *Progress Report* HD05631, "Intervention at Age Two and Three and Subsequent Intellective Performance," 2/14/75, Appendix A: Selection of Subjects in Harlem for the Early Training and School Performance Project (xerox) pp. 1-19.

11. Larry Schweinhart and David Weikart, "Can Preschool Make a Lasting Difference?" p. 3.

12. Willa Abelson, Edward Zigler and Cheryl DeBlasi, "Effects of a Four Year Follow Through Program on Economically Disadvantaged Children," Journal of Educational Psychology, Vol. 66, No. 55, 1974, pp. 757-758.

    Victoria Seitz, "Long-Term Effects of Intervention: A Longitudinal Investigation," (xerox).

13. Barry Guinagh and Ira Gordon, "School Performance as a Function of Early Stimulation," Institute paper presented at the American Psychological Association Annual Meeting, 1977, pp. 1-15.

14. Francis Palmer, "The Effects of Early Childhood Intervention," paper presented at the 1977 Annual Meeting of the American Association for the Advancement of Science, pp. 9-10.

Appendix B-4

RELATIONSHIP OF CURRENT RESULTS ON SPECIAL EDUCATION AND GRADE FAILURE TO "PRELIMINARY FINDINGS OF THE DEVELOPMENTAL CONTINUITY LONGITUDINAL STUDY"

The results reported in this paper differ slightly from those reported in "Preliminary Findings of the Developmental Continuity Longitudinal Study." The causes of these differences are as follows:

Gray: Notification of changes in the grade failure information on two subjects (one from "missing data" to "failure" and one from "no failures" to "failure").

Levenstein: The after-only control group (twenty-seven children) was excluded from this analysis for the reasons listed in Appendix B-1.

Miller: Four changes affected the Miller results:

- Speech and Hearing classes were excluded from the definition of special education for reasons explained in footnote 1. This resulted in the shift of six cases (five program and one control) from "special education" to "no special education."
- One child who had been listed as "special education (undefined)" had actually been in an advanced class and, hence, was reclassified as "no special education."
- An after-only control group (thirteen children) was excluded from the grade failure analysis. The group was not part of the original experimental design and was excluded for design reasons similar to those resulting in the exclusion of Levenstein's "after-only" group. The group had not been included in the special education analysis in "Preliminary Findings..." but had been included in the grade failure analysis.

- Data was received on two additional children. This affected only the grade failure analysis because the special education information was missing for both children.

Appendix B-5

GLOSSARY

ACYF	Administration for Children, Youth, and Families
ANOVA	Analysis of Variance
CD	Curriculum Demonstration Project (Weikart)
DARCEE	Demonstration and Research Center for Early Education developed at Peabody College (Miller)
ETP	Early Training Project (Gray)
ETS	Educational Testing Service
FOHV	Family Oriented Home Visitor Program (Gray)
GOAL	Game Oriented Activities for Learning (Karnes' Ameliorative)
ISP	Index of Social Position
NERA	Northeastern Educational Research Association
OCD	Office of Child Development (now ACYF - Administration for Children Youth and Families)
OEO	Office of Economic Opportunity
PCC	Parent-Child Center
PCDC	Parent-Child Development Center
PPVT	Peabody Picture Vocabulary Test
SES	Socio-economic Status
SPSS	Statistical Package for the Social Sciences
WAIS	Weschler Adult Intelligence Scale
WISC	Weschler Intelligence Scale for Children
WISC-R	Weschler Intelligence Scale for Children, Revised